GenCore version 5.1.6

ucleic - nucleic search, using sw model

on: July 22, 2003, 11:16:54 ; Search time 985 Seconds
(without alignments)
1211.386 Million cell updates/sec

le: U95626-C-AT-42723
fect score: 41
uence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

ring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

rched: 2054640 seqs, 14551402878 residues

al number of hits satisfying chosen parameters: 4109280

imum DB seq length: 0
mum DB seq length: 2000000000

t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	583	9	HSMCPA01	U80923 Human monoc
2	41	100.0	2082	9	AF068265	AF068265 Homo sapi
3	41	100.0	143068	6	AX335952	AX335952 Sequence
4	41	100.0	143068	9	HSU95626	U95626 Homo sapien
5	41	100.0	185437	9	AC098613	AC098613 Homo sapi
6	27.2	66.3	120793	2	AF322450	AF322450 Homo sapi
7	27.2	66.3	133544	9	AC100748	AC100748 Homo sapi
8	27.2	66.3	163035	2	AC006405	AC006405 Homo sapi
9	25.6	62.4	110394	9	AC003086	AC003086 Homo sapi
10	25.6	62.4	142082	2	AC093137	AC093137 Pan trogl
c 11	25.6	62.4	159365	9	AL392111	AL392111 Human DNA
c 12	25.6	62.4	174133	9	AC093183	AC093183 Homo sapi
13	25.6	62.4	176875	9	AC018638	AC018638 Homo sapi
14	25.6	62.4	178451	9	AL139396	AL139396 Human DNA
c 15	25.6	62.4	179110	9	AC011755	AC011755 Homo sapi
16	25.6	62.4	179503	9	AL445468	AL445468 Human DNA
17	25.6	62.4	183434	2	AC103924	AC103924 Homo sapi
18	25.6	62.4	190948	2	AC093138	AC093138 Pan trogl
c 19	25.6	62.4	210736	2	AC092792	AC092792 Homo sapi
20	25.2	61.5	132463	2	AC111775	AC111775 Rattus no
21	25.2	61.5	187803	10	AL596252	AL596252 Mouse DNA
22	25.2	61.5	188780	9	AC010896	AC010896 Homo sapi
23	25	61.0	68116	2	AC103970	AC103970 Homo sapi
c 24	25	61.0	73235	2	AC103725	AC103725 Homo sapi
25	25	61.0	78986	2	AC115626	AC115626 Homo sapi
26	25	61.0	82601	9	AL136971	AL136971 Human DNA
c 27	25	61.0	112204	9	AL136439	AL136439 Human DNA
28	25	61.0	113836	9	AL133174	AL133174 Human DNA
c 29	25	61.0	116845	9	AC008819	AC008819 Homo sapi
30	25	61.0	117583	9	AC034238	AC034238 Homo sapi
31	25	61.0	120689	9	CNS01RGS	AL159191 Human chr
c 32	25	61.0	133322	2	AC091845	AC091845 Homo sapi
c 33	25	61.0	142441	2	HSJ344H20	AL132710 Homo sapi
c 34	25	61.0	155428	9	AC009682	AC009682 Homo sapi
c 35	25	61.0	160979	2	AC055879	AC055879 Homo sapi
36	25	61.0	164809	9	AC128690	AC128690 Homo sapi
37	25	61.0	165146	2	AC011644	AC011644 Homo sapi
c 38	25	61.0	167163	2	AC013705	AC013705 Homo sapi
39	25	61.0	169337	2	AL356462	AL356462 Homo sapi
c 40	25	61.0	173834	9	AC108044	AC108044 Homo sapi
c 41	25	61.0	174701	9	AC007625	AC007625 Genomic s
c 42	25	61.0	182019	2	AC024388	AC024388 Homo sapi
c 43	25	61.0	187889	9	AC093008	AC093008 Homo sapi
c 44	25	61.0	193131	9	AL390766	AL390766 Human DNA
45	25	61.0	202750	2	AC116942	AC116942 Pan trogl

ALIGNMENTS

RESULT 1
HSMCPA01
LOCUS HSMCPA01 583 bp DNA linear PRI 24-JUL-1997
DEFINITION Human monocyte chemoattractant protein 1 receptor gene, 5' region.
ACCESSION U80923
VERSION U80923.1 GI:1773031
KEYWORDS .
SEGMENT 1 of 2
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Wong,L.M., Myers,S.J., Tsou,C.L., Gosling,J., Arai,H. and
Charo,I.F.

ITLE Organization and differential expression of the human monocyte chemoattractant protein 1 receptor gene. Evidence for the role of the carboxyl-terminal tail in receptor trafficking
OURNAL J. Biol. Chem. 272 (2), 1038-1045 (1997)
EDLINE 97150864
PUBMED 8995400
ERENCE 2 (bases 1 to 583)
UTHORS Myers,S.J. and Charo,I.F.
ITLE Direct Submission
OURNAL Submitted (04-DEC-1996) Pharmacology, Emory University, 1510 Clifton Road, Atlanta 30322, USA
TURES Location/Qualifiers
source 1. .583
/organism="Homo sapiens"
/db_xref="taxon:9606"
E COUNT 157 a 136 c 156 g 134 t
GIN

uery Match 100.0%; Score 41; DB 9; Length 583;
est Local Similarity 100.0%; Pred. No. 5.6e-05;
atches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||||||||||||||||||||||||||||||||||||||||||
Db 62 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 102

ULT 2
68265
US AF068265 2082 bp DNA linear PRI 20-APR-1999
INITION Homo sapiens monocyte chemoattractant protein 1 receptor (CCR2) gene promoter and mRNA, partial sequence.
ESSION AF068265
SION AF068265.1 GI:4587865
WORDS .
RCE Homo sapiens.
RGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ERENCE 1 (bases 1 to 2082)
UTHORS Yamamoto,K., Takeshima,H., Hamada,K., Nakao,M., Kino,T., Nishi,T., Kochi,M., Kuratsu,J., Yoshimura,T. and Ushio,Y.
ITLE Cloning and functional characterization of the 5'-flanking region of the human monocyte chemoattractant protein-1 receptor (CCR2) gene. Essential role of 5'-untranslated region in tissue-specific expression
OURNAL J. Biol. Chem. 274 (8), 4646-4654 (1999)
EDLINE 99143121
PUBMED 9988701
ERENCE 2 (bases 1 to 2082)
UTHORS Yamamoto,K., Takeshima,H., Hamada,K., Nakao,M., Kino,T., Nishi,T., Kochi,M., Kuratsu,J.-I., Yoshimura,T. and Ushio,Y.
ITLE Direct Submission
OURNAL Submitted (26-MAY-1998) Neurosurgery, Kumamoto University School of Medicine, Honjo, 1-1-1, Kumamoto 860-8556, Japan
TURES Location/Qualifiers
source 1. .2082
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
gene 1. .>2082
/gene="CCR2"
promoter 1. .1697
/gene="CCR2"
CAAT_signal 1570. .1574
/gene="CCR2"
TATA_signal 1666. .1672
/gene="CCR2"
mRNA 1698. .>2012
/gene="CCR2"
/product="monocyte chemoattractant protein 1 receptor"
5'UTR 1698. .2012
/gene="CCR2"
BASE COUNT 650 a 439 c 433 g 560 t
ORIGIN

Query Match 100.0%; Score 41; DB 9; Length 2082;
Best Local Similarity 100.0%; Pred. No. 4.3e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||||||||||||||||||||||||||||||||||||||||||
Db 1959 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 1999

RESULT 3
AX335952
LOCUS AX335952 143068 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6461 from Patent WO0194629.
ACCESSION AX335952
VERSION AX335952.1 GI:18126671
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G., Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature gene sets
JOURNAL Patent: WO 0194629-A 6461 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1. .143068
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 41194 a 30122 c 32403 g 39349 t
ORIGIN

Query Match 100.0%; Score 41; DB 6; Length 143068;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||||||||||||||||||||||||||||||||||||||||||
Db 42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743

RESULT 4
HSU95626
LOCUS HSU95626 143068 bp DNA linear PRI 16-MAY-1997
DEFINITION Homo sapiens ccr2b (ccr2), ccr2a (ccr2), ccr5 (ccr5) and ccr6 (ccr6) genes, complete cds, and lactoferrin (lactoferrin) gene, partial cds, complete sequence.
ACCESSION U95626
VERSION U95626.1 GI:2104517
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143068)
AUTHORS McCombie,W.R., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D., Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K., Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J., Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M., Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E., Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y., Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
TITLE Human BAC clone 110P12
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 143068)
AUTHORS McCombie,R.W., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D., Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K.,

```
        Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J.,
        Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M.,
        Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E.,
        Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y.,
        Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
TITLE   Direct Submission
JOURNAL Submitted (27-MAR-1997) Advanced Genome Sequence Analysis Course,
        Cold Spring Harbor Laboratory, 1Bungtown Rd., Cold Spring Harbor,
        NY 11724, USA
MENT    Regions with single-strand coverage are as follows:

          31434 - 31443   37900 - 37968   53303 - 53357
          59166 - 59206   63708 - 63998   65200 - 65335
          78605 - 78713   92135 - 92137  112377 - 112551
         112643 - 112778 134284 - 134309 134914 - 135019
         143046 - 144068.
TURES           Location/Qualifiers
   source       1. .143068
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="3"
                /clone="BAC 110P12"
   gene         46056. .49505
                /gene="ccr2"
                /note="confirmed by similarity to Human monocyte
                chemoattractant protein 1 receptor (ccr2) mRNA (Accession
                Number U80924), two alternatively spliced mRNAs."
   mRNA         join(46056. .47046,48255. .49505)
                /gene="ccr2"
                /product="ccr2a"
                /note="confirmed by similarity to Human monocyte
                chemoattractant protein 1 receptor (ccr2) alternatively
                spliced mRNA encoding A-form carboxyl tail, Accession
                Number U80924."
   mRNA         46056. .47997
                /gene="ccr2"
                /product="ccr2b"
                /note="confirmed by similarity to Human monocyte
                chemoattractant protein 1 receptor (ccr2) alternatively
                spliced mRNA encoding B-form carboxyl tail. Accession
                Number: U80924."
   CDS          join(46106. .47046,48255. .48438)
                /gene="ccr2"
                /note="confirmed by similarity to Human monocyte
                chemoattractant protein 1 receptor (ccr2) alternatively
                spliced A-form, Encoded by GenBank Accession Number
                U80924, gi 1168965"
                /codon_start=1
                /product="ccr2a"
                /protein_id="AAB57791.1"
                /db_xref="GI:2104518"
                /translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
                LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
                ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
                SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
                VICYSGILXTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
                CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRSLFHIALGCRIAPLQKPVCGG
                PGVRPGKNVKVTTQGLLDGRGKGKSIGRAPEASLQDKEGA"
   CDS          46106. .47188
                /gene="ccr2"
                /note="confirmed by similarity to Human monocyte
                chemoattractant protein 1 receptor (ccr2), Accession
                Number: 1168965"
                /codon_start=1
                /product="ccr2b"
                /protein_id="AAB57792.1"
                /db_xref="GI:2104519"
                /translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
                LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
                ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
                SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
                VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
                CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRRYLSVFFRKHITKRFCKQCPV
                FYRETVDGVTSTNTPSTGEQEVSAGL"
   gene         59531. .64785
                /gene="ccr5"
   mRNA         join(59531. .59573,61472. .64785)
                /gene="ccr5"
                /product="ccr5"
                /note="confirmed by similarity to Human cc chemokine
                receptor 5 (ccr5) mRNA. Accession number: U54994."
   CDS          61483. .62541
                /gene="ccr5"
                /note="confirmed by similarity to human CC chemokine
                receptor 5 (ccr5) protein, encoded by GenBank Accession
                Number U54994, gi 1457946"
                /codon_start=1
                /product="ccr5"
                /protein_id="AAB57793.1"
                /db_xref="GI:2104520"
                /translation="MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFG
                FVGNMLVILILINCKRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQ
                LLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFAS
                LPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVICYSGIL
                KTLLRCRNEKKRHRAVRLIFTIMIVYFLFWAPYNIVLLLNTFQEFFGLNNCSSSNRLD
                QAMQVTETLGMTHCCINPIIYAPVGEKFRNYLLVFFQKHIAKRFCKCCSIFQQEAPER
                ASSVYTRSTGEQEISVGL"
   mRNA         96634. .97683
                /product="ccr6"
                /note="Putative mRNA identified by homology to CCR5 mRNA."
                /evidence=not_experimental
   gene         96642. .97676
                /gene="ccr6"
   CDS          96642. .97676
                /gene="ccr6"
                /note="Identified as a gene by Grail Version 1.3c.
                Translated sequence exhibits 42% sequence identity to CCR5
                protein."
                /codon_start=1
                /evidence=not_experimental
                /product="ccr6"
                /protein_id="AAB57794.1"
                /db_xref="GI:2104521"
                /translation="MANYTLAPEDEYDVLIEGELESDEAEQCDKYDAQALSAQLVPSL
                CSAVFVIGVLDNLLVVLILVKYKGLKRVENIYLLNLAVSNLCFLLTLPFWAHAGGDPM
                CKILIGLYFVGLYSETFFNCLLTVQRYLVFLHKGNFFSARRRVPCGIITSVLAWVTAI
                LATLPEYVVYKPQMEDQKYKCAFSRTPFLPADETFWKHFLTLKMNISVLVLPLFIFTF
                LYVQMRKTLRFREQRYSLFKLVFAIMVVFLLMWAPYNIAFFLSTFKEHFSLSDCKSSY
                NLDKSVHITKLIATTHCCINPLLYAFLDGTFSKYLCRCFHLRSNTPLQPRGQSAQGTS
                REEPDHSTEV"
   gene         complement(124605. .>139255)
                /gene="lactoferrin"
   mRNA         complement(join(124605. .124816,126528. .126717,
                127884. .128068,130006. .130073,132023. .132164,
                133863. .134018,135022. .135075,135890. .135980,
                137445. .137599,138436. .138610,139077. .>139255))
                /gene="lactoferrin"
                /product="lactoferrin"
                /note="confirmed by similarity to lactoferrin mRNA,
                accession number M73700"
   CDS          complement(join(124782. .124816,126528. .126717,
                127884. .128068,130006. .130073,132023. .132164,
                133863. .134018,135022. .135075,135890. .135980,
                137445. .137599,138436. .138610,139077. .>139253))
                /gene="lactoferrin"
                /note="confirmed by similarity to lactoferrin protein,
                encoded by GenBank  Accession Number M73700, gi  186818"
                /codon_start=1
                /product="lactoferrin"
                /protein_id="AAB57795.1"
                /db_xref="GI:2104522"
                /translation="DLSDEAERDEYELLCPDNTRKPVDKFKDCHLARVPSHAVVARSV
                NGKEDAIWNLLRQAQEKFGKDKSPKFQLFGSPSGQKDLLFKDSAIGFSRVPPRIDSGL
                YLGSGYFTAIQNLRKSEEEVAARRARVVWCAVGEQELRKCNQWSGLSEGSVTCSSAST
                TEDCIALVLKGEADAMSLDGGYVYTAGKCGLVPVLAENYKSQQSSDPDPNCVDRPVEG
                YLAVAVVRRSDTSLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQTGSCKFDEYFSQS
```

```
                CAPGSDPRSNLCALCIGDEQGENKCVPNSNERYYGYTGAFRCLAENAGDVAFVKDVTV
                LQNTDGNNNDAWAKDLKLADFALLCLDGKRKPVTEARSCHLAMAPNHAVVSRMDKVER
                LKQVLLHQQAKFGRNGSDCPDKFCLFQSETKNLLFNDNTECLARLHGKTTYEKYLGPQ
                YVAGITNLKKCSTSPLLEACEFLRK"
BE COUNT    41194 a  30122 c  32403 g  39349 t
IGIN

Query Match           100.0%;  Score 41;  DB 9;  Length 143068;
Best Local Similarity  100.0%;  Pred. No. 1.9e-05;
Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
          |||||||||||||||||||||||||||||||||||||||||
    42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743


SULT 5
098613
CUS       AC098613              185437 bp    DNA     linear   PRI 01-AUG-2002
FINITION  Homo sapiens chromosome 3 clone RP11-24F11, complete sequence.
CESSION   AC098613
SION      AC098613.2  GI:22038607
YWORDS    HTG.
URCE      human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 185437)
AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
          Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
          and Haugen,E.D.
ITLE      Direct Submission
OURNAL    Unpublished
FERENCE   2  (bases 1 to 185437)
AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
ITLE      Direct Submission
OURNAL    Submitted (26-OCT-2001) Genome Center, University of Washington,
          Box 352145, Seattle, WA 98195, USA
ERENCE    3  (bases 1 to 185437)
AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
          Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
          and Haugen,E.D.
ITLE      Direct Submission
OURNAL    Submitted (01-AUG-2002) Genome Center, University of Washington,
          Box 352145, Seattle, WA 98195, USA
MMENT     On Aug 1, 2002 this sequence version replaced gi:16445164.
          ---------- Genome Center
              Center: University of Washington Genome Center
              Center Code: UWGC
              Web site: http://www.genome.washington.edu
              Contact: uwgchtgs@u.washington.edu
          ---------- Project Information
              Center project name: chr-3
              Center clone name: RP11-24F11 (bc0137)
          ---------- Summary Statistics
              Sequencing vector: plasmid; L08752; 100% of reads
              Chemistry: Dye-terminator ET; 93% of reads
              Chemistry: Dye-terminator Big Dye; 7% of reads
              Assembly program: Phrap; version 0.990319
              Consensus quality: 184860 bases at least Q40
              Consensus quality: 185398 bases at least Q30
              Consensus quality: 185435 bases at least Q20
              Insert size: 185437; sum-of-contigs
              Quality coverage: 7.6x in Q20 bases; sum-of-contigs
          ----------
          Overlapping Sequences:
              5': BAC-110P12 U95626, 111014-bp overlap
              3': RP11-509I21 (UWGC:bc0454) AC104304, 61294-bp overlap
          --------------------
          Sequence Quality Assessment:
           This entry has been annotated with sequence quality
           estimates computed by the Phrap assembly program.
           All manually edited bases have been reduced to quality zero.
           Quality levels above 40 are expected to have less than
           1 error in 10,000 bp.
           Base-by-base quality values are not generally visible from the
           GenBank flat file format but are available as part
           of this entry's ASN.1 file.

          This sequence was finished as follows unless otherwise noted:
          all regions were either double-stranded or sequenced with an
          alternate chemistry or covered by high quality data (i.e., Phred
          quality >= 30); an attempt was made to resolve all sequencing
          problems, such as compressions and repeats; all regions were
          covered by at least one plasmid subclone or more than one M13
          subclone; and the assembly was confirmed by restriction digest.
          --------------------
          Sequence Validation:
           This sequence has been validated by Multiple Complete Digest
           fingerprinting. Comparison of the experimentally derived digest
           fragments with sequence-predicted fragments is given below.
           The electronically-digested sequence consists of both insert and
           vector, in order to accurately represent the entire circular BAC.
           Small fragments below a variable cutoff (approximately 400-800 bp)
           are not resolved in the fingerprint and hence do not appear
           in the table. There are no significant remaining discrepancies
           between the experimental and predicted values. Uniquely ordered
           fragments are separated by dashed lines.
```

BglII		EcoRI		HindIII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
11786	12040	8696	8837	9531	9423
2067	2065	6	<800	6382	6501
5681	5720	1846	1824	512	<800
3716	3953	4052	3986	449	<800
875	897	1159	1130	6692	6501
910	897	54	<800	10449	10229
2169	2215	560	<800	953	1025
1705	1653	2287	2267	3124	3176
5763	5720	4905	4891	1054	1025
5844	5720	3049	3078	3985	3998
5061	5001	1377	1394	1100	1025
2625	2640	9903	9772	124	<800
725	<800	1022	1021	2671	2693
1173	1161	12606	12503	1948	1968
5747	5720	866	868	83	<800
9995	9684	9817	9772	7455	7474
999	999	3598	3601	1305	1267
4541	4503	452	<800	1047	1025
2950	2981	7549	7635	3299	3301
406	<800	2063	2075	5279	5237
3375	3602	5837	5858	6815	6926

13024	13045	13685	13328	435	<800
416	<800	4104	3986	2509	2576
10298	10140	1943	1928	2218	2172
1478	1452	3964	3986	3797	3802
287	<800	2758	2771	7758	7830
205	<800	167	<800	16192	16442
9428	9684	1307	1304	11020	10820
3809	3953	13306	13328	1978	1968
1275	1234	2423	2450	8741	8681
3351	3381	2243	2267	4795	4815
193	<800	1109	1130	14921	14985
334	<800	4131	3986	1896	1968
2546	2640	6383	6450	172	<800
9155	9107	332	<800	2091	2172
9190	9107	242	<800	2159	2172
352	<800	15571	15803	3573	3564
1157	1161	453	<800	5252	5237
3967	3953	1429	1394	4478	4428
1369	1334	655	<800	6166	6243
752	752	7881	7987	1565	1603
1624	1653	7533	7635	1645	1603
3258	3381	585	<800	573	<800
9624	9684	2998	2973	7986	7830
2841	2870	8796	8837	2313	2324
2979	3069	1331	1304	49	<800
3898	3953	7102	7161	1250	1267
4830	4804			2290	2324
2418	2459			2056	1968
6198	6066				
534	<800				
3985	3953				
1184	1161				
63	<800				

TURES Location/Qualifiers
source 1. .185437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-24F11"

Query Match 100.0%; Score 41; DB 9; Length 185437;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||||||||||||||||||||||||||||||||||||||||
Db 10653 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 10693

RESULT 6
AF322450
LOCUS AF322450 120793 bp DNA linear HTG 06-DEC-2000
DEFINITION Homo sapiens chromosome 17 clone BAC407I21 map 17p13.3, ***
SEQUENCING IN PROGRESS ***, 28 unordered pieces.
ACCESSION AF322450
VERSION AF322450.1 GI:11559855
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 120793)
AUTHORS Zhao,X.T., He,M., Wan,D.F., Ye,Y., Qin,W.X., Huang,Y., Zuo,L. and
Gu,J.R.
TITLE Gene clone on human chromosome 17p13.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 120793)
AUTHORS Zhao,X.T., He,M., Wan,D.F., Ye,Y., Qin,W.X., Huang,Y., Zuo,L. and
Gu,J.R.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2000) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
200032, P. R. China
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 608: contig of 608 bp in length
* gap of unknown length
* 609 1282: contig of 674 bp in length
* gap of unknown length
* 1283 1733: contig of 451 bp in length
* gap of unknown length
* 1734 2435: contig of 702 bp in length
* gap of unknown length
* 2436 2895: contig of 460 bp in length
* gap of unknown length
* 2896 3643: contig of 748 bp in length
* gap of unknown length
* 3644 4149: contig of 506 bp in length
* gap of unknown length
* 4150 4908: contig of 759 bp in length
* gap of unknown length
* 4909 5741: contig of 833 bp in length
* gap of unknown length
* 5742 6777: contig of 1036 bp in length
* gap of unknown length
* 6778 7766: contig of 989 bp in length
* gap of unknown length
* 7767 9183: contig of 1417 bp in length
* gap of unknown length
* 9184 9931: contig of 748 bp in length
* gap of unknown length
* 9932 12274: contig of 2343 bp in length
* gap of unknown length
* 12275 16059: contig of 3785 bp in length
* gap of unknown length
* 16060 19969: contig of 3910 bp in length
* gap of unknown length

```
          *    19970    22905: contig of 2936 bp in length
          *                    gap of unknown length
          *    22906    28210: contig of 5305 bp in length
          *                    gap of unknown length
          *    28211    34020: contig of 5810 bp in length
          *                    gap of unknown length
          *    34021    41704: contig of 7684 bp in length
          *                    gap of unknown length
          *    41705    48302: contig of 6598 bp in length
          *                    gap of unknown length
          *    48303    56789: contig of 8487 bp in length
          *                    gap of unknown length
          *    56790    65996: contig of 9207 bp in length
          *                    gap of unknown length
          *    65997    74434: contig of 8438 bp in length
          *                    gap of unknown length
          *    74435    83698: contig of 9264 bp in length
          *                    gap of unknown length
          *    83699    95956: contig of 12258 bp in length
          *                    gap of unknown length
          *    95957   107953: contig of 11997 bp in length
          *                    gap of unknown length
          *   107954   120793: contig of 12840 bp in length.
ATURES             Location/Qualifiers
   source          1. .120793
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="17"
                   /map="17p13.3"
                   /clone="BAC407I21"
SE COUNT    28674 a  31051 c  31727 g  29285 t     56 others
IGIN

Query Match             66.3%;  Score 27.2;  DB 2;  Length 120793;
Best Local Similarity   80.0%;  Pred. No. 3.2;
Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            |||||||| |||||||||| ||    |||| ||| |||| |
         49 CAGTTGCTCAGAAGCCTGAGATGAGAGGATTGCTTGAGCC 88


SULT 7
100748
OCUS       AC100748              133544 bp    DNA     linear   PRI 01-JUN-2002
EFINITION  Homo sapiens chromosome 17, clone CTD-2231E3, complete sequence.
CCESSION   AC100748
ERSION     AC100748.2  GI:21306865
EYWORDS    HTG.
OURCE      human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE   1  (bases 1 to 133544)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
ITLE       Homo sapiens chromosome 17, clone CTD-2231E3
OURNAL     Unpublished
EFERENCE   2  (bases 1 to 133544)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
           Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
           Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
           Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
           MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
           McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
           Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
           Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
           Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
           Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
           Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE    Direct Submission
  JOURNAL  Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE  3  (bases 1 to 133544)
  AUTHORS  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
           Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
           Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
           Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
           Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
           Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
           Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
           Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
           Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
           Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE    Direct Submission
  JOURNAL  Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Jun 1, 2002 this sequence version replaced gi:17048115.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
           -------------- Project Information
               Center project name: L17901
               Center clone name: 2231_E_3
           --------------.
FEATURES             Location/Qualifiers
     source          1. .133544
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2231E3"
                     /clone_lib="CITD1 Human BAC"
     repeat_region   2. .92
                     /rpt_family="AluY"
     repeat_region   642. .705
                     /rpt_family="L2"
     repeat_region   954. .1030
                     /rpt_family="AluSq/x"
     repeat_region   1033. .1159
                     /rpt_family="AluY"
     repeat_region   1162. .1470
                     /rpt_family="AluY"
     repeat_region   1474. .1578
                     /rpt_family="L2"
     repeat_region   1811. .2124
                     /rpt_family="AluSq"
     repeat_region   2227. .2356
                     /rpt_family="FLAM_C"
```

```
repeat_region   2359. .2664
                /rpt_family="AluY"
repeat_region   2666. .2972
                /rpt_family="AluSc"
repeat_region   3119. .3305
                /rpt_family="GA-rich"
repeat_region   complement(3411. .3646)
                /rpt_family="AluSx"
repeat_region   3707. .3742
                /rpt_family="(TAA)n"
repeat_region   complement(3747. .3951)
                /rpt_family="MIR"
repeat_region   4038. .4315
                /rpt_family="L1MC5"
repeat_region   4317. .4343
                /rpt_family="AT_rich"
repeat_region   complement(4344. .4729)
                /rpt_family="L1PA5"
repeat_region   complement(4730. .5037)
                /rpt_family="AluSc"
repeat_region   complement(5038. .5114)
                /rpt_family="L1PA5"
repeat_region   complement(5121. .5409)
                /rpt_family="AluSx"
repeat_region   5530. .6202
                /rpt_family="L2"
repeat_region   complement(6244. .6561)
                /rpt_family="AluSx"
repeat_region   complement(6571. .6882)
                /rpt_family="AluSq"
repeat_region   7169. .7189
                /rpt_family="GC_rich"
repeat_region   7208. .7292
                /rpt_family="(CGG)n"
repeat_region   7309. .7392
                /rpt_family="(CGGG)n"
repeat_region   7488. .7672
                /rpt_family="(CGGGG)n"
repeat_region   8823. .8853
                /rpt_family="(TG)n"
unsure          11173. .11391
                /note="single clone coverage"
repeat_region   11182. .11203
                /rpt_family="GC_rich"
unsure          11191. .11196
                /note="<30 qual single clone coverage"
unsure          11210. .11217
                /note="<30 qual single clone coverage"
unsure          11342. .11347
                /note="<30 qual single clone coverage"
unsure          11374. .11379
                /note="<30 qual single clone coverage"
repeat_region   complement(11725. .12024)
                /rpt_family="AluSx"
repeat_region   complement(13025. .13151)
                /rpt_family="AluJo"
repeat_region   15438. .15582
                /rpt_family="AluSg/x"
repeat_region   16432. .16574
                /rpt_family="FLAM_C"
repeat_region   complement(17703. .17810)
                /rpt_family="L1MB3"
repeat_region   18378. .18699
                /rpt_family="AluJo"
repeat_region   19047. .19117
                /rpt_family="L2"
repeat_region   19228. .19445
                /rpt_family="AluJo"
repeat_region   19727. .19806
                /rpt_family="L2"
repeat_region   19999. .20309
                /rpt_family="AluSx"
repeat_region   20356. .20652
                /rpt_family="AluY"
repeat_region   20768. .21077
                /rpt_family="AluSg"
repeat_region   complement(21954. .22261)
                /rpt_family="AluSx"
repeat_region   complement(22299. .22581)
                /rpt_family="AluSx"
repeat_region   complement(22597. .22895)
                /rpt_family="AluJb"
repeat_region   22956. .22976
                /rpt_family="(TTTTG)n"
repeat_region   23745. .24055
                /rpt_family="AluSq"
repeat_region   complement(24948. .25089)
                /rpt_family="AluSx"
repeat_region   complement(25144. .25317)
                /rpt_family="AluSx"
repeat_region   27682. .27757
                /rpt_family="MIR"
repeat_region   complement(27758. .28055)
                /rpt_family="AluSq"
repeat_region   28056. .28168
                /rpt_family="MIR"
repeat_region   28853. .29073
                /rpt_family="AluJo"

  Query Match             66.3%;  Score 27.2;  DB 9;  Length 133544;
  Best Local Similarity   80.0%;  Pred. No. 3.1;
  Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||||||| |||||||||| ||   |||| ||| |||| |
Db     103679 CAGTTGCTCAGAAGCCTGAGATGAGAGGATTGCTTGAGCC 103718


RESULT 8
AC006405
LOCUS       AC006405              163035 bp    DNA     linear   HTG 21-JAN-1999
DEFINITION  Homo sapiens chromosome 17 clone hRPK.107_N_19 map 17, ***
            SEQUENCING IN PROGRESS ***, 12 unordered pieces.
ACCESSION   AC006405
VERSION     AC006405.1  GI:4165361
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163035)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.107_N_19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163035)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 21, 1999 this sequence version replaced gi:4159875.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
```

```
          http://ftp.genome.washington.edu/RM/RepeatMasker.html.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     8815: contig of 8815 bp in length
          *                   gap of unknown length
          *    8816    12285: contig of 3470 bp in length
          *                   gap of unknown length
          *   12286    40956: contig of 28671 bp in length
          *                   gap of unknown length
          *   40957    64393: contig of 23437 bp in length
          *                   gap of unknown length
          *   64394    78975: contig of 14582 bp in length
          *                   gap of unknown length
          *   78976    96637: contig of 17662 bp in length
          *                   gap of unknown length
          *   96638   105967: contig of 9330 bp in length
          *                   gap of unknown length
          *  105968   145832: contig of 39865 bp in length
          *                   gap of unknown length
          *  145833   147041: contig of 1209 bp in length
          *                   gap of unknown length
          *  147042   150498: contig of 3457 bp in length
          *                   gap of unknown length
          *  150499   160979: contig of 10481 bp in length
          *                   gap of unknown length
          *  160980   163035: contig of 2056 bp in length.
TURES             Location/Qualifiers
   source         1. .163035
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="17"
                  /map="17"
                  /clone="hRPK.107_N_19"
                  /clone_lib="RPCI-11 human BAC library"
E COUNT    38384 a  43845 c  42661 g  37996 t    149 others
GIN

uery Match             66.3%;  Score 27.2;  DB 2;  Length 163035;
est Local Similarity   80.0%;  Pred. No. 3;
atches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||||||| |||||||||| ||   |||| ||| |||| |
      97969 CAGTTGCTCAGAAGCCTGAGATGAGAGGATTGCTTGAGCC 98008


ULT 9
003086
CUS       AC003086              110394 bp    DNA     linear   PRI 21-DEC-1999
INITION   Homo sapiens BAC clone CTB-104F4 from 7q21-q22, complete sequence.
CESSION   AC003086
SION      AC003086.1  GI:2588618
WORDS     HTG.
URCE      Homo sapiens.
RGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 110394)
AUTHORS   Sulston,J.E. and Waterston,R.
ITLE      Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
 PUBMED   9847074
FERENCE   2  (bases 1 to 110394)
AUTHORS   Bradshaw,H., Graves,T. and Biewald,T.
ITLE      The sequence of Homo sapiens BAC clone CTB-104F4
JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 110394)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 110394)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 110394)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
              Center project name: H_RG104F04
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center.
            For additional information about the map position of this sequence,
            see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone CTB-104F4 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is CTB-5F13, 200 bp overlap; the
            clone sequenced to the right is CTA-293F11, 200 bp overlap.  Actual
            start of this clone is at base position 78804 of CTB-5F13; actual
            end is at 110200 of CTB-104F4.

            This clone contains polymorphic bases with CTA-293F11.
FEATURES              Location/Qualifiers
     source           1. .110394
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /map="7q21-q22"
                      /clone="CTB-104F4"
                      /clone_lib="CITB-978SK-B"
     repeat_region    131. .175
                      /rpt_family="L1"
     repeat_region    1395. .1416
                      /rpt_family="L1"
```

```
repeat_region   1466. .1756
                /rpt_family="ALU"
repeat_region   2000. .2036
                /rpt_family="L1"
repeat_region   5535. .5590
                /rpt_family="L1"
repeat_region   complement(6440. .6522)
                /rpt_family="L1"
misc_feature    6804. .7069
                /note="match to EST T81349 (NID:g704234) yd27c10.s1"
misc_feature    complement(7736. .8097)
                /note="match to EST T81524 (NID:g704531) yd27c10.r1"
repeat_region   8413. .8499
                /rpt_family="L1"
repeat_region   8686. .9935
                /rpt_family="L1"
repeat_region   complement(9944. .10235)
                /rpt_family="ALU"
repeat_region   10236. .10689
                /rpt_family="L1"
repeat_region   10690. .10814
                /rpt_family="ALU"
repeat_region   10816. .12134
                /rpt_family="L1"
repeat_region   12149. .13515
                /rpt_family="L1"
repeat_region   complement(12574. .12982)
                /rpt_family="L1"
repeat_region   complement(13467. .13560)
                /rpt_family="L1"
repeat_region   13759. .13809
                /rpt_family="L1"
repeat_region   13941. .14230
                /rpt_family="ALU"
repeat_region   14315. .14376
                /rpt_family="L1"
repeat_region   15011. .15113
                /rpt_family="L1"
repeat_region   16643. .16668
                /rpt_family="L1"
repeat_region   18883. .19174
                /rpt_family="ALU"
repeat_region   19257. .19767
                /rpt_family="L1"
repeat_region   19768. .20060
                /rpt_family="ALU"
repeat_region   20080. .20596
                /rpt_family="L1"
repeat_region   22793. .22822
                /rpt_family="L1"
repeat_region   complement(23139. .23254)
                /rpt_family="ALU"
misc_feature    23255. .23722
                /note="match to EST N32155 (NID:g1152554) yy23b12.s1"
misc_feature    23255. .23458
                /note="match to EST R19544 (NID:g773154) yg26f05.r1"
repeat_region   complement(23798. .23850)
                /rpt_family="L1"
misc_feature    complement(24191. .24304)
                /note="match to EST R43896 (NID:g821774) yg26f05.s1"
repeat_region   complement(26269. .26552)
                /rpt_family="ALU"
repeat_region   26990. .27191
                /rpt_family="L1"
repeat_region   27233. .27284
                /rpt_family="L1"
misc_feature    complement(27804. .28176)
                /note="match to EST R83407 (NID:g928284) yp82d02.r1"
gene            complement(29090. .35579)
                /gene="WUGSC:H_RG104F04.1"
CDS             complement(join(29090. .30260,35551. .35579))
                /gene="WUGSC:H_RG104F04.1"
                /note="match to Y09615 (PID:g1707507) (NID:g1707506);
                H_RG104F04.1"
                /codon_start=1
                /product="mitochondrial transcription terminator factor"
                /protein_id="AAB83941.1"
                /db_xref="GI:2588619"
                /translation="MQSLSLGQTSISKGLNYLTIMAPGNLWHMRNNPLFGSRCWMTRF
                SAENIFKSVSFRLFGVKCHNTDSEPLKNEDLLKNLLTMGVDIDMARKRQPGVFHRMIT
                NEQDLKMFLLSKGASKEVIASIISRYPRAITRTPENLSKRWDLWRKIVTSDLEIVNIL
                ERSPESFFRSNNNLNLENNIKFLYSVGLTRKCLCRLLTNAPRTFSNSLDLNKQMVEFL
                QAAGLSLGHNDPADFVRKIIFKNPFILIQSTKRVKANIEFLRSTFNLNSEELLVLICG
                PGAEILDLSNDYARRSYANIKEKLFSLGCTEEEVQKFVLSYPDVIFLAEKKFNDKIDC
                LMEENISISQIIENPRVLDSSISTLKSRIKELVNAGCNLSTLNITLLSWSKKRYEAKL
                KKLSRFA"
misc_feature    complement(29284. .29492)
                /gene="WUGSC:H_RG104F04.1"
                /note="match to EST R45772 (NID:g804496)"
misc_feature    complement(30026. .30266)
                /gene="WUGSC:H_RG104F04.1"
                /note="match to EST D60352 (NID:g961991)"
repeat_region   complement(30754. .31049)
                /rpt_family="ALU"
repeat_region   34049. .34338
                /rpt_family="ALU"
repeat_region   34376. .34648
                /rpt_family="ALU"
repeat_region   34693. .34795
                /rpt_family="L1"
repeat_region   37172. .37462
                /rpt_family="ALU"
repeat_region   complement(37511. .37826)
                /rpt_family="ALU"
repeat_region   38964. .39005
                /rpt_family="ALU"
repeat_region   40696. .40733
                /rpt_family="L1"
```

Query Match 62.4%; Score 25.6; DB 9; Length 110394;
Best Local Similarity 77.5%; Pred. No. 13;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| |||| ||||| |||| || |||||||| |||| |
Db      34215 CAGCTGCTCAGAAGGCTGAGGTAGGAGGACTGCTTGAGCC 34254
```

```
RESULT 10
AC093137
LOCUS       AC093137              142082 bp    DNA     linear   HTG 12-JUN-2002
DEFINITION  Pan troglodytes clone RP43-144H14, WORKING DRAFT SEQUENCE, 4
            ordered pieces.
ACCESSION   AC093137
VERSION     AC093137.2  GI:21392484
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 142082)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142082)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (11-AUG-2001) NIH Intramural Sequencing Center, 8717
          Grovemont Circle, Gaithersburg, MD 20877, USA
FERENCE   3  (bases 1 to 142082)
AUTHORS   Green,E.D.
TITLE     Direct Submission
JOURNAL   Submitted (12-JUN-2002) NIH Intramural Sequencing Center, 8717
          Grovemont Circle, Gaithersburg, MD 20877, USA
MMENT     On Jun 12, 2002 this sequence version replaced gi:15148127.
          -------------- Genome Center
               Center: NIH Intramural Sequencing Center
               Center code: NISC
               Web site: http://www.nisc.nih.gov
               Contact: nisc_zoo@nhgri.nih.gov
          -------------- Project Information
               Center project name: anl
               Center clone name: 144H14

          The sequence data in this record represents an 'enhanced'
          version of a Phase 2 submission. Specifically, the indicated
          order and orientation of each sequence contig has been
          established using one or more of the following: read-pair
          data from individual subclones, overlaps with neighboring
          clones, alignment with available reference sequence (e.g.,
          human), and/or confirmation by PCR testing. In addition,
          the sequence assembly is based on at least 8X average
          coverage in Q20 bases and has been reviewed to rule out
          gross misassemblies, the low-quality ends of sequence
          contigs have been trimmed away, and each base is associated
          with a Phrap-derived quality score.
          -------------- Summary Statistics
               Sequencing vector: plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 141565 bases at least Q40
               Consensus quality: 141698 bases at least Q30
               Consensus quality: 141749 bases at least Q20
               Insert size: 113000; agarose-fp
               Insert size: 100000; pulse-field-gel
               Insert size: 141782; sum-of-contigs
               Quality coverage: 13.44x in Q20 bases; agarose-fp
               Quality coverage: 15.18x in Q20 bases; pulse-field-gel
               Quality coverage: 10.71x in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    47102: contig of 47102 bp in length
          *    47103    47202: gap of unknown length
          *    47203    86175: contig of 38973 bp in length
          *    86176    86275: gap of unknown length
          *    86276   118675: contig of 32400 bp in length
          *   118676   118775: gap of unknown length
          *   118776   142082: contig of 23307 bp in length.
ATURES              Location/Qualifiers
     source         1. .142082
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="RP43-144H14"
                    /clone_lib="RP43"
     misc_feature   1. .47102
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:left"
     misc_feature   12236. .142082
                    /note="clone overlaps with GenBank Accession Number
                    AC093138 clone RP43-154A20 (center project name anm)"
     misc_feature   47203. .86175
                    /note="assembly_fragment"
     misc_feature   86276. .118675
                    /note="assembly_fragment"
     misc_feature   118776. .142082
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:right"
     misc_feature   132713. .142082
                    /note="clone overlaps with GenBank Accession Number
                    AC093189 clone RP43-17P3 (center project name ann)"
BASE COUNT    42473 a  27490 c  27247 g  44571 t    301 others
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 2;  Length 142082;
  Best Local Similarity   77.5%;  Pred. No. 12;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| |||| ||||| |||| || |||||||| |||| |
Db      38952 CAGCTGCTCAGAAGACTGAGGTAGGAGGACTGCTTGAGCC 38991


RESULT 11
AL392111/c
LOCUS       AL392111              159365 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-487I5 on chromosome 10, complete
            sequence.
ACCESSION   AL392111
VERSION     AL392111.12  GI:16972970
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159365)
  AUTHORS   Bray-Allen,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15808188.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-487I5 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-487I5 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-487I5 is at 159365 in this
            sequence. The true left end of clone RP11-375G3 is at 149681 in
            this sequence. The true right end of clone RP11-77G23 is at 4000 in
```

```
        this sequence.
TURES             Location/Qualifiers
   source         1. .159365
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="10"
                  /clone="RP11-487I5"
                  /clone_lib="RPCI-11.2"
SE COUNT    41838 a  35376 c  36999 g  45152 t
IGIN

Query Match             62.4%;  Score 25.6;  DB 9;  Length 159365;
Best Local Similarity   77.5%;  Pred. No. 12;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || ||||| ||||   |  ||||||||||||||
     103353 CAGCTACTCAGAAGGCTGATGCAGGAGGACTGCCTGAGAC 103314


SULT 12
093183/c
OCUS        AC093183              174133 bp    DNA     linear   PRI 29-MAR-2002
DEFINITION  Homo sapiens chromosome 7 clone RP11-786I1, complete sequence.
ACCESSION   AC093183
VERSION     AC093183.3  GI:19807864
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174133)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 174133)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (14-AUG-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 174133)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:17998609.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-7
                Center clone name: RP11-786I1 (djs736)
            ------------- Summary Statistics
                Sequencing vector: plasmid; 56% of reads
                Sequencing vector: plasmid; L08752; 44% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 174068 bases at least Q40
                Consensus quality: 174122 bases at least Q30
                Consensus quality: 174133 bases at least Q20
                Insert size: 174133; sum-of-contigs
                Quality coverage: 7.7x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  RP11-212P7 AC090114 32817-bp clone overlap
                3':  RP11-274B21 (UWGC:djs180) AC018638 115239-bp clone overlap
            --------------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
            This sequence has been validated by Multiple Complete Digest
            fingerprinting. Comparison of the experimentally derived digest
            fragments with sequence-predicted fragments is given below.
            The electronically-digested sequence consists of both insert and
            vector, in order to accurately represent the entire circular BAC.
            Small fragments below a variable cutoff (approximately 400-800 bp)
            are not resolved in the fingerprint and hence do not appear
            in the table. There are no significant remaining discrepancies
            between the experimental and predicted values. Uniquely ordered
            fragments are separated by dashed lines.
```

EcoRI		HindIII		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8841	808	799	3746	3589
6	<800	6382	6526	2067	2136
4135	4154	512	<800	9001	9123
9341	9257	449	<800	4328	4355
962	970	6297	6216	1041	1054
3828	3866	14754	14447	2308	2467
8990	8841	116	<800	931	890
1284	1261	1855	1981	10550	10542
4271	4259	7682	8065	3920	4229
4644	4639	5348	5312	630	<800
1874	1837	3865	3811	6869	6855
5493	5514	3113	3148	404	<800
8821	8841	242	<800	1878	1877
3234	3211	1512	1518	8398	8427
292	<800	7335	7311	1322	1326
1725	1714	2808	2839	687	<800
1059	1061	579	<800	7972	8003
2055	2073	10626	10446	5685	5677
8123	8154	980	986	976	890
1808	1837	8983	9103	6594	6629

1511	1511	135	<800	3862	3947
4909	4926	1983	2088	7222	7284
4590	4639	3085	3148	2355	2467
1889	1837	9737	9812	11057	10958
4919	4926	4041	4014	883	890
18867	18932	4428	4505	101	<800
11516	11472	224	<800	869	890
5914	5942	2126	2088	3181	3224
8939	8841	8331	8538	10120	10208
92	<800	236	<800	323	<800
3796	3866	3925	4014	38	<800
1363	1351	4908	4977	1864	1877
2077	2073	1215	1217	13497	13350
2559	2594	20880	21062	835	890
3523	3535	257	<800	2149	2136
8210	8154	516	<800	6576	6629
6569	6514	4509	4505	20896	20939
96	<800	2105	2088	3378	3224
1617	1632	4687	4694	546	<800
4908	4926	1822	1844	9754	9704
2070	2073	5512	5541	3553	3420
		91	<800	466	<800
		1664	1646		
		1151	1127		
		1847	1844		
		6048	6024		
		3123	3148		

```
ATURES             Location/Qualifiers
   source          1. .174133
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="7"
                   /clone="RP11-786I1"
                   /clone_lib="RPCI human BAC library 11"
   misc_feature    90126. .90268
                   /note="Single subclone region"
   misc_feature    148430. .148575
                   /note="Single subclone region"
   misc_feature    148949. .149235
                   /note="Single subclone region"
   misc_feature    174037. .174133
                   /note="Single subclone region"
SE COUNT    45264 a  38500 c  37886 g  52483 t
IGIN
```

```
  Query Match            62.4%;  Score 25.6;  DB 9;  Length 174133;
  Best Local Similarity  77.5%;  Pred. No. 12;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| ||||||| |||||||| |||| |
Db      86529 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 86490


RESULT 13
AC018638
LOCUS       AC018638              176875 bp    DNA     linear   PRI 27-JAN-2001
DEFINITION  Homo sapiens chromosome 7 clone RP11-274B21, complete sequence.
ACCESSION   AC018638
VERSION     AC018638.5  GI:12583819
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176875)
  AUTHORS   Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
            Olson,M.V.
  TITLE     Large-scale Mapping and Sequencing of Human Chromosome 7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176875)
  AUTHORS   Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1999) Human Genome Center, University of
            Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 176875)
  AUTHORS   Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
            Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Jan 27, 2001 this sequence version replaced gi:9719754.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: HsaChr7
                Center clone name: RP11-274B21 (djs180)
            ------------- Summary Statistics
                Sequencing vector: M13; X02513; 100% of reads
                Chemistry: Dye-primer-amersham; 99% of reads
                Chemistry: Dye-terminator ET-amersham; 1% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 176700 bases at least Q40
                Consensus quality: 176852 bases at least Q30
                Consensus quality: 176872 bases at least Q20
                Insert size: 186780; 5.8% error; agarose-fp
                Insert size: 176875; sum-of-contigs
                Quality coverage: 8.57x in Q20 bases; agarose-fp
                Quality coverage: 9.05x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5': RP11-66F23 (UWGC:djs178) AC024952, 72585-bp clone overlap
                3': RP11-20F2 (UWGC:djs173)
            -------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
```

all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI		NsiI		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8266	20007	20170	2589	2562
6	<800	2189	2217	2067	2086
1059	1053	2222	2217	7030	7221
2055	2050	2633	2650	976	977
8123	8266	334	<800	6594	6600
1808	1818	226	<800	3861	3902
1510	1594	2433	2455	7222	7221
4909	4919	9830	9904	2355	2377
4590	4670	1941	1898	11058	10963
1889	1818	16117	16057	883	869
4920	4919	3548	3575	101	<800
18867	18788	11972	12004	869	869
11516	11684	2958	2952	3181	3204
5915	5970	6864	6910	10120	10205
8939	9018	5934	6024	323	<800
92	<800	917	919	38	<800
3796	3863	2121	2109	1864	1875
1363	1320	152	<800	13497	13351
2077	2050	3177	3223	835	869
2559	2583	2646	2650	2150	2086
3524	3517	407	<800	6576	6600
8210	8266	1039	1040	20897	21438
6569	6560	1797	1791	3378	3384
96	<800	3908	3915	546	<800
1617	1689	12688	12552	9754	9698
4908	4670	5248	5223	3553	3552
2070	2050	1815	1791	466	<800
2757	2795	534	<800	1777	1758
2687	2681	1753	1741	1930	1875
6157	6146	3091	3090	687	<800
4425	4464	13176	13152	6313	6264
4631	4670	3691	3697	2278	2285
1477	1453	6125	6024	4274	4304
1717	1818	3694	3697	1202	1189
4810	4670	6063	6024	2628	2654
1089	1053	2204	2217	158	<800
2871	2890	478	<800	2351	2377
13319	13333	17790	17691	4869	4776
12452	12521	1852	1846	3202	3204
1459	1453			2047	2086
4040	4012			21915	21438
2102	2086				
1902	1875				
2994	2992				
162	<800				

```
FEATURES             Location/Qualifiers
     source          1. .176875
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-274B21"
                     /sex="male"
                     /tissue_type="blood"
                     /clone_lib="RPCI human BAC library 11"
     STS             5530. .5596
                     /standard_name="sWSS3556"
     misc_feature    9059. .9141
                     /note="Single subclone region"
     STS             12255. .12543
                     /standard_name="sWSS1325"
     STS             complement(17643. .17844)
                     /standard_name="sWSS3779"
     STS             complement(100521. .100620)
                     /standard_name="sWSS2227"
     STS             complement(135945. .136020)
                     /standard_name="sWSS3427"
BASE COUNT    50911 a  38997 c  39737 g  47230 t
ORIGIN

  Query Match           62.4%;  Score 25.6;  DB 9;  Length 176875;
  Best Local Similarity  77.5%;  Pred. No. 12;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   ||||||||  |||| |
Db     149241 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 149280
```

```
SULT 14
139396
CUS       AL139396              178451 bp    DNA     linear   PRI 28-FEB-2002
 INITION  Human DNA sequence from clone RP11-258C19 on chromosome
          Xp11.21-11.23, complete sequence.
CESSION   AL139396
SION      AL139396.18  GI:19068211
YWORDS    HTG.
URCE      human.
RGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ERENCE   1  (bases 1 to 178451)
AUTHORS   Chapman,J.
ITLE      Direct Submission
OURNAL    Submitted (28-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
MMENT     On Mar 1, 2002 this sequence version replaced gi:10862725.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome X, constructed by the Sanger Centre Chromosome X Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          This sequence is the entire insert of clone RP11-258C19 This
          sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. RP11-258C19 is from
          the library RPCI-11.1 constructed by the group of Pieter de Jong.
          For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6.
ATURES              Location/Qualifiers
   source           1. .178451
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /map="p11.21-11.23"
                    /clone="RP11-258C19"
                    /clone_lib="RPCI-11.1"
   misc_feature     178349. .178451
                    /note="Single clone region. Assembly confirmed by
                    restriction digest data"
SE COUNT    48101 a  42307 c  40594 g  47449 t
IGIN

Query Match             62.4%;  Score 25.6;  DB 9;  Length 178451;
Best Local Similarity   77.5%;  Pred. No. 12;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| |||| || |||||||   |  |||||||||||||| |
      61898 CAGCTGCTTAGGAGCCTGAGGCAGGAGGACTGCCTGAGCC 61937


SULT 15
011755/c
CUS       AC011755              179110 bp    DNA     linear   PRI 07-NOV-2001
```

```
DEFINITION  Homo sapiens BAC clone RP11-575M22 from 2, complete sequence.
ACCESSION   AC011755
VERSION     AC011755.7  GI:14140349
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179110)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 179110)
  AUTHORS   Martinka,S., Abbott,A. and Boyer,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-575M22
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 179110)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 179110)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 179110)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 17, 2001 this sequence version replaced gi:13270817.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_NH0575M22
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
```

```
        VECTOR:  pBACe3.6
        NEIGHBORING SEQUENCE INFORMATION:
        The clone sequenced to the left is RP11-93I3, 200 bp overlap; the
        clone sequenced to the right is RP11-449L24, 200 bp overlap.
        Actual start of this clone is at base position 195 of RP11-575M22;
        actual end is at base position 178916 of RP11-575M22.

        RP11-575M22 contains single stranded regions from 78912 to 79086
        and 123395 to 123463.
ATURES              Location/Qualifiers
   source           1. .179110
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="2"
                    /map="2"
                    /clone="RP11-575M22"
                    /clone_lib="RPCI-11"
   repeat_region    105. .127
                    /rpt_family="AT_rich"
   repeat_region    258. .338
                    /rpt_family="MER2_type"
   repeat_region    467. .866
                    /rpt_family="MaLR"
   repeat_region    1757. .1831
                    /rpt_family="MIR"
   repeat_region    1917. .2043
                    /rpt_family="MIR"
   repeat_region    3551. .3856
                    /rpt_family="L1"
   repeat_region    3918. .4005
                    /rpt_family="L2"
   repeat_region    4062. .4136
                    /rpt_family="L1"
   repeat_region    4390. .4571
                    /rpt_family="MER1_type"
   repeat_region    5831. .6116
                    /rpt_family="L1"
   repeat_region    7659. .7725
                    /rpt_family="(TTCA)n"
   repeat_region    7743. .8144
                    /rpt_family="L2"
   repeat_region    8337. .8482
                    /rpt_family="MIR"
   repeat_region    8869. .9083
                    /rpt_family="MER1_type"
   repeat_region    9971. .10095
                    /rpt_family="MIR"
   repeat_region    10111. .10224
                    /rpt_family="L2"
   repeat_region    10309. .10605
                    /rpt_family="Alu"
   repeat_region    11779. .11927
                    /rpt_family="L2"
   repeat_region    12527. .12779
                    /rpt_family="MER1_type"
   repeat_region    13379. .13629
                    /rpt_family="MER1_type"
   repeat_region    14028. .14161
                    /rpt_family="Alu"
   repeat_region    14192. .14619
                    /rpt_family="MaLR"
   repeat_region    14827. .15204
                    /rpt_family="MaLR"
   repeat_region    15380. .15499
                    /rpt_family="L2"
   repeat_region    15548. .15849
                    /rpt_family="Alu"
   repeat_region    16029. .16444
                    /rpt_family="MaLR"
   repeat_region    16445. .16630
                    /rpt_family="MER2_type"
   repeat_region    16957. .17105
                    /rpt_family="MIR"
   repeat_region    18513. .19133
                    /rpt_family="MaLR"
   repeat_region    19308. .19523
                    /rpt_family="MaLR"
   repeat_region    19557. .19649
                    /rpt_family="MaLR"
   repeat_region    19767. .19831
                    /rpt_family="Mariner"
   repeat_region    20151. .20482
                    /rpt_family="Alu"
   misc_feature     21962. .22347
                    /note="similar to  EST AW751954 (NID:g7666990)"
   misc_feature     22130. .22493
                    /note="similar to  EST T07269 (NID:g318418)"
   repeat_region    23745. .23932
                    /rpt_family="MER1_type"
   repeat_region    24295. .24451
                    /rpt_family="MIR"
   repeat_region    24818. .25577
                    /rpt_family="L1"
   repeat_region    26071. .26387
                    /rpt_family="Alu"
   repeat_region    26687. .26732
                    /rpt_family="(CA)n"
   repeat_region    26806. .26954
                    /rpt_family="MIR"
   repeat_region    27219. .27363
                    /rpt_family="MIR"
   repeat_region    27454. .27722
                    /rpt_family="Alu"
   repeat_region    29988. .30259
                    /rpt_family="Alu"
   repeat_region    30749. .30844
                    /rpt_family="MIR"
   repeat_region    31573. .31690
                    /rpt_family="L2"
   repeat_region    31893. .32044
                    /rpt_family="CR1"
   repeat_region    32410. .32433
                    /rpt_family="(TGAA)n"
   repeat_region    32544. .32924
                    /rpt_family="L1"
   misc_feature     33116. .33387
                    /note="similar to  EST BE082071 (NID:g8472367)"
   repeat_region    33494. .33516
                    /rpt_family="AT_rich"
   repeat_region    34711. .35101
                    /rpt_family="MaLR"
   repeat_region    36087. .36139
                    /rpt_family="MIR"
   repeat_region    36140. .36428
                    /rpt_family="Alu"
   repeat_region    36429. .36472

  Query Match             62.4%;  Score 25.6;  DB 9;  Length 179110;
  Best Local Similarity   77.5%;  Pred. No. 12;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||||| || || || |||||||| |||| ||| |||| |
Db     149688 CAGTTACTCAGGAGGCTGACATAAAAGGATTGCTTGAGCC 149649


Search completed: July 22, 2003, 13:43:32
Job time : 992 secs
```

GenCore version 5.1.6

nucleic - nucleic search, using sw model

on: July 22, 2003, 11:16:53 ; Search time 170.5 Seconds
(without alignments)
541.536 Million cell updates/sec

le: U95626-C-AT-42723
fect score: 41
quence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

rched: 2185239 seqs, 1125999159 residues

al number of hits satisfying chosen parameters: 4370478

imum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	143068	21	AAF21105	Human low adenosin
2	41	100.0	143068	21	AAF21272	Human low adenosin
3	41	100.0	143068	21	AAA34983	Human adenosine re
4	41	100.0	143068	21	AAA35150	Human adenosine re
5	41	100.0	143068	24	ABL68124	Ovary cancer relat
6	41	100.0	149412	21	AAA35151	Human adenosine re
7	41	100.0	152740	21	AAF21273	Human low adenosin
8	25.8	62.9	633	22	AAL14736	Human breast cance
9	25.6	62.4	1001	22	AAI81867	Human polynucleoti
10	25.2	61.5	604	22	AAL20925	Human breast cance
11	25.2	61.5	604	22	AAL23598	Human breast cance
c 12	24	58.5	349	22	ABA14274	Human nervous syst
c 13	24	58.5	397	22	AAI91318	Human polynucleoti
c 14	24	58.5	935	24	AAS16521	Human cDNA encodin
c 15	24	58.5	992	23	AAS75085	DNA encoding novel
c 16	24	58.5	1624	22	ABA15771	Human nervous syst
17	24	58.5	1624	22	AAK73388	Human immune/haema
18	24	58.5	3565	16	AAQ84051	Sequence encoding
c 19	24	58.5	4749	22	AAH18519	Human cDNA sequenc
20	24	58.5	10528	24	ABL32334	Human immune syste
c 21	24	58.5	17335	23	ABK42393	Genomic sequence #
c 22	24	58.5	19882	23	ABK42394	Genomic sequence #
23	23.4	57.1	413	22	AAI81647	Human polynucleoti
c 24	23.4	57.1	10528	24	ABL32335	Human immune syste
c 25	23.4	57.1	38374	24	ABN96966	Gene #3464 used to
c 26	23.4	57.1	38374	24	ABL68363	Kidney cancer rela
c 27	23.4	57.1	38374	24	ABL68364	Kidney cancer rela
c 28	23.4	57.1	38374	24	ABL68824	Kidney cancer rela
29	23	56.1	178896	24	ABQ88146	Human osteoblast d
c 30	22.8	55.6	1912	20	AAV83767	Rhodococcus strain
c 31	22.6	55.1	908	23	AAS84558	DNA encoding novel
c 32	22.6	55.1	75384	22	AAK85590	Human immune/haema
c 33	22.4	54.6	329	22	AAI91783	Human polynucleoti
34	22.4	54.6	360	22	AAI82165	Human polynucleoti
35	22.4	54.6	387	22	AAK67876	Human immune/haema
c 36	22.4	54.6	395	22	AAI82659	Human polynucleoti
c 37	22.4	54.6	441	23	ABV56385	Human prostate exp
38	22.4	54.6	445	22	AAK86122	Human immune/haema
39	22.4	54.6	445	22	AAK86123	Human immune/haema
40	22.4	54.6	445	22	AAK86124	Human immune/haema
c 41	22.4	54.6	450	22	AAS32604	Human genomic DNA
c 42	22.4	54.6	564	23	AAS71886	DNA encoding novel
43	22.4	54.6	574	23	ABV30511	Human prostate exp
44	22.4	54.6	574	23	ABV39484	Human prostate exp
45	22.4	54.6	799	23	ABV09335	Human prostate exp

ALIGNMENTS

RESULT 1
AAF21105
ID AAF21105 standard; DNA; 143068 BP.
XX
AC AAF21105;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2672.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.

(NYCE/) NYCE J W.

Nyce JW;

WPI; 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not trigger adenosine receptors during metabolism, useful e.g. for treating cancers and respiratory obstructions -

Disclosure; Page 924-957; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

```
Query Match           100.0%;  Score 41;  DB 21;  Length 143068;
Best Local Similarity  100.0%;  Pred. No. 1.6e-06;
Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            |||||||||||||||||||||||||||||||||||||||||
      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

SULT 2
F21272
AAF21272 standard; DNA; 143068 BP.

AAF21272;

14-MAR-2001 (first entry)

Human low adenosine antisense oligonucleotide related sequence #2839.

Low adenosine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1186-1219; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

```
  Query Match           100.0%;  Score 41;  DB 21;  Length 143068;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-06;
  Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||||||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

RESULT 3
AAA34983
ID AAA34983 standard; DNA; 143068 BP.
XX
AC AAA34983;
XX
DT 28-JUL-2000 (first entry)
XX

Human adenosine receptor related polynucleotide SEQ ID NO:2672.

Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

Homo sapiens.

WO200009525-A2.

24-FEB-2000.

03-AUG-1999; 99WO-US17712.

03-AUG-1998; 98US-0095212.

(UYEC-) UNIV EAST CAROLINA.

Nyce JW;

WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

Disclosure; Page 851-882; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 143068 BP; 41194 A; 30122 C; 32402 G; 39350 T; 0 other;

uery Match 100.0%; Score 41; DB 21; Length 143068;
cst Local Similarity 100.0%; Pred. No. 1.6e-06;
atches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
        1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
          |||||||||||||||||||||||||||||||||||||||||
    42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

ULT 4
35150

ID AAA35150 standard; DNA; 143068 BP.
XX
AC AAA35150;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:24.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1106-1138; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 143068 BP; 41194 A; 30126 C; 32402 G; 39346 T; 0 other;

Query Match 100.0%; Score 41; DB 21; Length 143068;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
```

```
                 ||||||||||||||||||||||||||||||||||||||||
     42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743

RESULT 5
ABL68124
ABL68124 standard; DNA; 143068 BP.

ABL68124;

15-MAY-2002  (first entry)

Ovary cancer related gene sequence SEQ ID NO:6461.

Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
gene; ds.

Homo sapiens.

WO200194629-A2.

13-DEC-2001.

30-MAY-2001; 2001WO-US10838.

05-JUN-2000; 2000US-209473P.
05-JUN-2000; 2000US-209531P.
18-SEP-2000; 2000US-233133P.
18-SEP-2000; 2000US-233617P.
20-SEP-2000; 2000US-234009P.
20-SEP-2000; 2000US-234034P.
20-SEP-2000; 2000US-234052P.
22-SEP-2000; 2000US-234509P.
22-SEP-2000; 2000US-234567P.
25-SEP-2000; 2000US-234923P.
25-SEP-2000; 2000US-234924P.
25-SEP-2000; 2000US-235077P.
25-SEP-2000; 2000US-235082P.
25-SEP-2000; 2000US-235134P.
25-SEP-2000; 2000US-235280P.
26-SEP-2000; 2000US-235637P.
26-SEP-2000; 2000US-235638P.
27-SEP-2000; 2000US-235711P.
27-SEP-2000; 2000US-235720P.
27-SEP-2000; 2000US-235840P.
27-SEP-2000; 2000US-235863P.
28-SEP-2000; 2000US-236028P.
28-SEP-2000; 2000US-236032P.
28-SEP-2000; 2000US-236033P.
28-SEP-2000; 2000US-236034P.
28-SEP-2000; 2000US-236109P.
28-SEP-2000; 2000US-236111P.
29-SEP-2000; 2000US-236842P.
29-SEP-2000; 2000US-236891P.
02-OCT-2000; 2000US-237172P.
02-OCT-2000; 2000US-237173P.
02-OCT-2000; 2000US-237278P.
02-OCT-2000; 2000US-237294P.
02-OCT-2000; 2000US-237295P.
02-OCT-2000; 2000US-237316P.
03-OCT-2000; 2000US-237425P.
03-OCT-2000; 2000US-237598P.
03-OCT-2000; 2000US-237604P.
03-OCT-2000; 2000US-237606P.
03-OCT-2000; 2000US-237608P.
01-NOV-2000; 2000US-244867P.
01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.

Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;

PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set  -
XX
PS   Claim 1; SEQ ID 6461; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

  Query Match             100.0%;  Score 41;  DB 24;  Length 143068;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||||||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743


RESULT 6
AAA35151
ID   AAA35151 standard; DNA; 149412 BP.
XX
AC   AAA35151;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:25.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
```

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

Disclosure; Page 1138-1171; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 149412 BP; 43049 A; 31388 C; 33852 G; 41123 T; 0 other;

```
Query Match             100.0%;  Score 41;  DB 21;  Length 149412;
Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

        1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
          |||||||||||||||||||||||||||||||||||||||||
    49047 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 49087
```

RESULT 7
AAF21273
AAF21273 standard; DNA; 152740 BP.

AAF21273;

14-MAR-2001 (first entry)

Human low adenosine antisense oligonucleotide related sequence #2840.

Low adenosine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

Homo sapiens.

WO200062736-A2.

26-OCT-2000.

24-MAR-2000; 2000WO-US08020.

XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1219-1254; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 152740 BP; 44169 A; 32023 C; 34549 G; 41999 T; 0 other;

```
  Query Match             100.0%;  Score 41;  DB 21;  Length 152740;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   41;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||||||||||||||||||||||||
Db      49047 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 49087
```

RESULT 8
AAL14736
ID AAL14736 standard; cDNA; 633 BP.
XX
AC AAL14736;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 7193.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX

19-JUL-2001.

10-JAN-2001; 2001WO-US00798.

14-JAN-2000; 2000US-0176077.
14-MAR-2000; 2000US-0189167.
24-MAR-2000; 2000US-0192099.
29-MAR-2000; 2000US-0193480.
15-MAY-2000; 2000US-0205230.
09-JUN-2000; 2000US-0211315.
25-JUL-2000; 2000US-0220534.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Lillie J, Xu Y, Wang Y, Steinmann K;

WPI; 2001-451856/48.

New peptide useful as a marker for the diagnosis of breast cancer -

Claim 1; Page 1300; 3695pp; English.

The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

Sequence 633 BP; 223 A; 92 C; 178 G; 136 T; 4 other;

```
Query Match             62.9%;  Score 25.8;  DB 22;  Length 633;
Best Local Similarity   78.9%;  Pred. No. 0.75;
Matches   30;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGA 38
            |||| |||| |||||||||||  ||  | ||| |||
        522 TCAGCTGCTCAGAAGCCTGACGTANGAGAATTGCTTGA 559
```

SULT 9
T81867/c
AAI81867 standard; cDNA; 1001 BP.

AAI81867;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 1927.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorders; arthritis; inflammation; ss.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US04927.

28-FEB-2000; 2000US-0515126.
18-MAY-2000; 2000US-0577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-514838/56.
DR P-PSDB; AAO01936.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 1927; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1001 BP; 258 A; 272 C; 239 G; 229 T; 3 other;

```
  Query Match             62.4%;  Score 25.6;  DB 22;  Length 1001;
  Best Local Similarity   77.5%;  Pred. No. 1;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   ||||||| |||| |
Db        908 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 869
```

RESULT 10
AAL20925
ID AAL20925 standard; cDNA; 604 BP.
XX
AC AAL20925;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 13382.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 2374; 3695pp; English.
XX

The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

Sequence 604 BP; 224 A; 82 C; 168 G; 130 T; 0 other;

Query Match 61.5%; Score 25.2; DB 22; Length 604;
Best Local Similarity 78.9%; Pred. No. 1.3;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGA 38
              |||| |||| |||||||||| ||  || | ||| |||
Db        495 TCAGCTGCTCAGAAGCCTGACGTAAGAGAATTGCTTGA 532
```

RESULT 11
AAL23598
AAL23598 standard; cDNA; 604 BP.

AAL23598;

07-DEC-2001 (first entry)

Human breast cancer expressed polynucleotide 16055.

Human; breast cancer; cell marker; cytostatic; ss.

Homo sapiens.

WO200151628-A2.

19-JUL-2001.

10-JAN-2001; 2001WO-US00798.

14-JAN-2000; 2000US-0176077.
14-MAR-2000; 2000US-0189167.
24-MAR-2000; 2000US-0192099.
29-MAR-2000; 2000US-0193480.
15-MAY-2000; 2000US-0205230.
09-JUN-2000; 2000US-0211315.
25-JUL-2000; 2000US-0220534.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Lillie J, Xu Y, Wang Y, Steinmann K;

WPI; 2001-451856/48.

New peptide useful as a marker for the diagnosis of breast cancer -

Claim 1; Page 2930; 3695pp; English.

The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

Sequence 604 BP; 224 A; 82 C; 168 G; 130 T; 0 other;

Query Match 61.5%; Score 25.2; DB 22; Length 604;
Best Local Similarity 78.9%; Pred. No. 1.3;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGA 38
              |||| |||| |||||||||| ||  || | ||| |||
Db        495 TCAGCTGCTCAGAAGCCTGACGTAAGAGAATTGCTTGA 532
```

RESULT 12
ABA14274/c
ID ABA14274 standard; cDNA; 349 BP.
XX
AC ABA14274;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 3281.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
20-OCT-2000; 2000US-0242221.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
DR P-PSDB; ABB17948.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 1; SEQ ID NO 3281; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 349 BP; 98 A; 96 C; 76 G; 76 T; 3 other;

Query Match 58.5%; Score 24; DB 22; Length 349;
Best Local Similarity 75.0%; Pred. No. 3.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||| | ||| |||| ||||||| |||| ||| |||| |
Db 164 CAGSTACTGGGAAGKCTGACATGGTAGGATTGCTTGAGCC 125

RESULT 13
AAI91318/c
ID AAI91318 standard; cDNA; 397 BP.
XX
AC AAI91318;
XX

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 11378.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorders; arthritis; inflammation; ss.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US04927.

28-FEB-2000; 2000US-0515126.
18-MAY-2000; 2000US-0577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-514838/56.
P-PSDB; AAO11387.

Isolated nucleic acids and polypeptides, useful for preventing diagnosing and treating e.g. leukaemia, inflammation and immune disorders -

Claim 1; SEQ ID NO 11378; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AAI79941-AAI93841) and the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 397 BP; 98 A; 110 C; 98 G; 91 T; 0 other;

Query Match 58.5%; Score 24; DB 22; Length 397;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   |||| ||| |||| |
Db        113 CAGCTACTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 74
```

SULT 14
S16521/c
AAS16521 standard; cDNA; 935 BP.

AAS16521;

14-FEB-2002 (first entry)

Human cDNA encoding serine protease 8.

Human; ss; serine protease 8; cytostatic; virucidal; immunomodulatory; antiinflammatory; haemostatic; malignant tumour; haemopathy; HIV; human immunodeficiency virus infection; immunological disease; inflammation.

XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 125..376
FT /*tag= a
FT /product= "Serine protease 8"
XX
PN WO200171003-A1.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-CN00404.
XX
PR 24-MAR-2000; 2000CN-0115084.
XX
PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2002-049036/06.
DR P-PSDB; AAU10876.
XX
PT Human serine protease 8 and encoded polynucleotide, used in diagnosis
PT and treatment of malignant tumours, haemopathy, human immunodeficiency
PT virus infection, immunological diseases and inflammation -
XX
PS Claim 6; Page 28; 33pp; Chinese.
XX
CC The invention relates to an isolated polypeptide of human serine
CC protease 8, its fragment, analogue or derivative, its encoding
CC polynucleotide, a recombinant vector (or viral vector) expressing the
CC protein, a host cell transformed with the vector, an antibody raised
CC against the protein, and mimics or regulators (ant/agonists) of the
CC protein's activity or expression. The protein and polynucleotide are used
CC in diagnosis and treatment of malignant tumours, haemopathy, human
CC immunodeficiency virus (HIV) infection, immunological diseases and
CC various inflammations. The present sequence encodes the human
CC serine protease 8.
XX
SQ Sequence 935 BP; 232 A; 267 C; 230 G; 206 T; 0 other;

Query Match 58.5%; Score 24; DB 24; Length 935;
Best Local Similarity 75.0%; Pred. No. 4.4;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   |||| ||| |||| |
Db        782 CAGCTACTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 743
```

RESULT 15
AAS75085/c
ID AAS75085 standard; cDNA; 992 BP.
XX
AC AAS75085;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10889.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.
P-PSDB; ABG10898.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 1; SEQ ID No 10889; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 992 BP; 226 A; 295 C; 276 G; 195 T; 0 other;

Query Match 58.5%; Score 24; DB 23; Length 992;
Best Local Similarity 75.0%; Pred. No. 4.5;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| |||||||   |||| ||| |||| |
        315 CAGCTACTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 276
```

arch completed: July 22, 2003, 12:24:07
b time : 174.5 secs

GenCore version 5.1.6

nucleic - nucleic search, using sw model

on: July 22, 2003, 11:16:51 ; Search time 36 Seconds
(without alignments)
349.271 Million cell updates/sec

le: U95626-C-AT-42723
fect score: 41
quence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

rched: 441362 seqs, 153338381 residues

al number of hits satisfying chosen parameters: 882724

imum DB seq length: 0
ximum DB seq length: 2000000000

t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description
1	24	58.5	3565	1	US-08-578-649-3	Sequence 3, Appli
2	22.4	54.6	5892	3	US-08-755-587-27	Sequence 27, Appl
3	22.4	54.6	15328	2	US-08-888-497-33	Sequence 33, Appl
4	22.4	54.6	15328	4	US-09-362-230-33	Sequence 33, Appl
5	22.4	54.6	15328	5	PCT-US94-07926-33	Sequence 33, Appl
6	22.4	54.6	87350	3	US-08-781-891-79	Sequence 79, Appl
7	22.4	54.6	87543	4	US-09-791-211-3	Sequence 3, Appli
8	22	53.7	5162	2	US-08-916-917-13	Sequence 13, Appl
9	22	53.7	5162	3	US-09-225-170-13	Sequence 13, Appl
10	21.8	53.2	393	4	US-09-574-141A-97	Sequence 97, Appl
11	21.8	53.2	399	4	US-09-574-141A-94	Sequence 94, Appl
12	21.8	53.2	907	3	US-09-081-320-39	Sequence 39, Appl
13	21.8	53.2	907	4	US-09-574-141A-39	Sequence 39, Appl
14	21.8	53.2	907	4	US-09-707-780-39	Sequence 39, Appl
15	21.8	53.2	3224	4	US-08-965-729A-2	Sequence 2, Appli
16	21.8	53.2	6485	3	US-09-081-320-2	Sequence 2, Appli
17	21.8	53.2	6485	4	US-09-574-141A-2	Sequence 2, Appli
18	21.8	53.2	6485	4	US-09-707-780-2	Sequence 2, Appli
19	21.8	53.2	8743	3	US-09-081-320-1	Sequence 1, Appli
20	21.8	53.2	8743	4	US-09-574-141A-1	Sequence 1, Appli
21	21.8	53.2	8743	4	US-09-707-780-1	Sequence 1, Appli
22	21.4	52.2	1866	3	US-08-909-742-1	Sequence 1, Appli
23	21.4	52.2	1866	4	US-09-412-289-1	Sequence 1, Appli
24	21.4	52.2	43950	4	US-09-735-934A-3	Sequence 3, Appli
25	21.2	51.7	12141	4	US-09-488-671-10	Sequence 10, Appl
26	21	51.2	1845	4	US-08-887-534A-22	Sequence 22, Appl
27	20.8	50.7	75	4	US-09-461-697-151	Sequence 151, App
28	20.8	50.7	399	4	US-09-574-141A-91	Sequence 91, Appl
29	20.8	50.7	542	4	US-09-461-697-136	Sequence 136, App
30	20.8	50.7	1284	3	US-09-081-320-35	Sequence 35, Appl
31	20.8	50.7	1284	4	US-09-574-141A-35	Sequence 35, Appl
32	20.8	50.7	1284	4	US-09-707-780-35	Sequence 35, Appl
c 33	20.8	50.7	1320	1	US-08-599-252-84	Sequence 84, Appl
c 34	20.8	50.7	1320	1	US-08-436-074-57	Sequence 57, Appl
c 35	20.8	50.7	1320	5	PCT-US96-06352-84	Sequence 84, Appl
c 36	20.8	50.7	1320	5	PCT-US96-06583-84	Sequence 84, Appl
37	20.8	50.7	1400	4	US-09-245-281-40	Sequence 40, Appl
38	20.8	50.7	1400	4	US-09-207-359B-40	Sequence 40, Appl
39	20.8	50.7	8133	4	US-09-659-791A-10	Sequence 10, Appl
40	20.8	50.7	8353	3	US-08-611-587-1	Sequence 1, Appli
41	20.8	50.7	32042	4	US-09-245-281-44	Sequence 44, Appl
42	20.8	50.7	162450	4	US-09-345-882-1	Sequence 1, Appli
43	20.8	50.7	176373	3	US-09-128-155-17	Sequence 17, Appl
44	20.6	50.2	3441	4	US-09-026-033-17	Sequence 17, Appl
45	20.2	49.3	399	4	US-09-574-141A-90	Sequence 90, Appl

ALIGNMENTS

RESULT 1
US-08-578-649-3
; Sequence 3, Application US/08578649
; Patent No. 5770366
; GENERAL INFORMATION:
; APPLICANT: Ulrich Bogdan
; APPLICANT: Reinhard Buttner
; APPLICANT: Brigitte Kaluza
; TITLE OF INVENTION: MELANOMA-INHIBITING PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/578,649
; FILING DATE: 29-July-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 43 24 247.2
; FILING DATE: 20-July-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Andrew L. Tiajoloff
; REGISTRATION NUMBER: 31,575
; REFERENCE/DOCKET NUMBER: BOER 1035-PFF/ALT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3565 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1378..1449
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1378..1504
; FEATURE:

```
     NAME/KEY:  exon
     LOCATION:  1586..1719
   FEATURE:
     NAME/KEY:  exon
     LOCATION:  2804..2914
   FEATURE:
     NAME/KEY:  exon
     LOCATION:  3232..3252
   FEATURE:
     NAME/KEY:  -
     LOCATION:  one-of(2216)
     OTHER INFORMATION:  /note= "N in position 2216
     OTHER INFORMATION:  denotes an indefinite number ans sequence of
     OTHER INFORMATION:  nucleotides"
-08-578-649-3

Query Match             58.5%;  Score 24;  DB 1;  Length 3565;
Best Local Similarity   75.0%;  Pred. No. 0.49;
Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||||| || || |||||||  |   |||| |||||||| |
        519 CAGTTACTCAGGAGCCTGAGGTGGGAGGATTGCCTGAGTC 558


SULT 2
-08-755-587-27/c
Sequence 27, Application US/08755587
Patent No. 6045997
 GENERAL INFORMATION:
   APPLICANT:  Futreal, Phillip A
   APPLICANT:  Wooster, Richard F
   APPLICANT:  Ashworth, Alan
   APPLICANT:  Stratton, Michael R
   TITLE OF INVENTION:  Materials and methods relating to the
   TITLE OF INVENTION:  identification and sequencing of the BRCA2 cancer
   TITLE OF INVENTION:  susceptibility gene and uses thereof.
   NUMBER OF SEQUENCES:  222
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Bell Seltzer Park & Gibson
     STREET:  310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
     CITY:  Raleigh
     STATE:  NC
     COUNTRY:  USA
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/755,587
     FILING DATE:  25-NOV-1996
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  GB 9523959.6
     FILING DATE:  23-NOV-1995
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  GB 9525555.0
     FILING DATE:  14-DEC-1995
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  GB 9617961.9
     FILING DATE:  28-AUG-1996
   ATTORNEY/AGENT INFORMATION:
     NAME:  Kenneth D Sibley
     REGISTRATION NUMBER:  31,665
     REFERENCE/DOCKET NUMBER:  5405-135
 INFORMATION FOR SEQ ID NO:  27:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  5892 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  double
     TOPOLOGY:  linear
   MOLECULE TYPE:  DNA (genomic)
   FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  483..5412
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  481..5412
US-08-755-587-27

  Query Match             54.6%;  Score 22.4;  DB 3;  Length 5892;
  Best Local Similarity   72.5%;  Pred. No. 2.7;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||||| || || ||  |||  | | |||||||| |||| |
Db        231 CAGTTACTCAGGAGGGTGAGGTGCAAGGACTGCTTGAGCC 192


RESULT 3
US-08-888-497-33
; Sequence 33, Application US/08888497
; Patent No. 5972677
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,497
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,405
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2498
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-888-497-33

  Query Match             54.6%;  Score 22.4;  DB 2;  Length 15328;
  Best Local Similarity   72.5%;  Pred. No. 3.4;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | || ||||| |||| |||  ||||   |||||| |
Db       8735 CAGCTACTCAGAAGGCTGAGATAGGAGGATCACCTGAGCC 8774
```

```
SULT 4
-09-362-230-33
Sequence 33, Application US/09362230
Patent No. 6352849
 GENERAL INFORMATION:
   APPLICANT:  Tischfield, Jay A.
   APPLICANT:  Seilhamer, Jeffrey J.
   TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
   TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
   TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
   TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
   NUMBER OF SEQUENCES:  44
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
     ADDRESSEE:  Russell PA
     STREET:  200 East Broward Boulevard
     CITY:  Fort Lauderdale
     STATE:  FL
     COUNTRY:  USA
     ZIP:  33301
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/09/362,230
     FILING DATE:
     CLASSIFICATION:
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/888,497
     FILING DATE:
     APPLICATION NUMBER:  US 08/097,354
     FILING DATE:  26-JUL-1993
   ATTORNEY/AGENT INFORMATION:
     NAME:  Manso, Peter J.
     REGISTRATION NUMBER:  32,264
     REFERENCE/DOCKET NUMBER:  IN21044-5
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  305-527-2498
     TELEFAX:  305-764-4996
 INFORMATION FOR SEQ ID NO:  33:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  15328 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  cDNA
-09-362-230-33

Query Match             54.6%;  Score 22.4;  DB 4;  Length 15328;
Best Local Similarity   72.5%;  Pred. No. 3.4;
Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || ||||| |||| |||  ||||   |||||| |
       8735 CAGCTACTCAGAAGGCTGAGATAGGAGGATCACCTGAGCC 8774


SULT 5
1-US94-07926-33
Sequence 33, Application PC/TUS9407926
 GENERAL INFORMATION:
   APPLICANT:  Tischfield, Jay A.
   APPLICANT:  Seilhamer, Jeffrey J.
   TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
   TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
   TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
   TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
   NUMBER OF SEQUENCES:  44
   CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07926
;      FILING DATE:  15-JUL-1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2498
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US94-07926-33

  Query Match             54.6%;  Score 22.4;  DB 5;  Length 15328;
  Best Local Similarity   72.5%;  Pred. No. 3.4;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | || ||||| |||| |||  ||||   |||||| |
Db       8735 CAGCTACTCAGAAGGCTGAGATAGGAGGATCACCTGAGCC 8774


RESULT 6
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
```

ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
-08-781-891-79

Query Match 54.6%; Score 22.4; DB 3; Length 87350;
Best Local Similarity 72.5%; Pred. No. 5.5;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || || |  |||| |||  |||||||| |||| |
      63492 CAGCTACTCAGGATGCTGAGATAAGAGGACTGCTTGAGCC 63453
```

SULT 7
-09-791-211-3/c
Sequence 3, Application US/09791211
Patent No. 6448080
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
FILE REFERENCE: RTS-0205
CURRENT APPLICATION NUMBER: US/09/791,211
CURRENT FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 90
SEQ ID NO 3
LENGTH: 87543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 7421
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 7427
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 11609
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 12605
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 12742
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29370
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29422
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29979
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29980
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29981
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 30136
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30140
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31205
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31206
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31592
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 33095
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 33160
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 34066
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 34072
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 36816
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 39020
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 42164
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 42459
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46808
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46823
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 46826
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 47291
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 52786
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 52787
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 53384
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 54684
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59215
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59235
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 59242
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 63290
; OTHER INFORMATION: unknown

```
  NAME/KEY: unsure
  LOCATION: 66614
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 68660
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 68697
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 68718
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 68733
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 68739
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 69785
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 79134
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 79198
  OTHER INFORMATION: unknown
  NAME/KEY: unsure
  LOCATION: 86336
  OTHER INFORMATION: unknown
  OTHER INFORMATION:
-09-791-211-3

Query Match             54.6%;  Score 22.4;  DB 4;  Length 87543;
Best Local Similarity   72.5%;  Pred. No. 5.5;
Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || || | |||| ||| |||||||| |||| |
      63685 CAGCTACTCAGGATGCTGAGATAAGAGGACTGCTTGAGCC 63646


SULT 8
-08-916-917-13
Sequence 13, Application US/08916917
Patent No. 5856132
 GENERAL INFORMATION:
   APPLICANT:  Stephens, Len
   APPLICANT:  Hawkins, Phillip Thomas
   APPLICANT:  Braselmann, Sylvia
   TITLE OF INVENTION:  G-BETA-GAMMA REGULATED
   TITLE OF INVENTION:  PHOSPHATIDYLINOSITOL-3' KINASE
   NUMBER OF SEQUENCES:  14
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Pennie & Edmonds, LLP
     STREET:  1155 Avenue of the Americas
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10036-2811
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  Windows
     SOFTWARE:  FastSEQ for Windows Version 2.0b
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/916,917
     FILING DATE:  15-AUG-1997
     CLASSIFICATION:  435
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/672,211
     FILING DATE:  27-JUN-1996
   ATTORNEY/AGENT INFORMATION:
;      NAME:  Abrams, Samuel B
;      REGISTRATION NUMBER:  30,605
;      REFERENCE/DOCKET NUMBER:  8549-0006-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-493-4935
;      TELEFAX:  650-493-5556
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5162 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-916-917-13

  Query Match             53.7%;  Score 22;  DB 2;  Length 5162;
  Best Local Similarity   70.0%;  Pred. No. 3.8;
  Matches   28;  Conservative    1;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | || || || ||||:  |  |||| |||||||| |
Db       5012 CAGGTACTCAGGAGGCTGASGCAGGAGGATTGCCTGAGCC 5051


RESULT 9
US-09-225-170-13
; Sequence 13, Application US/09225170
; Patent No. 6017763
;  GENERAL INFORMATION:
;    APPLICANT:  Stephens, Len
;    APPLICANT:  Hawkins, Phillip Thomas
;    APPLICANT:  Braselmann, Sylvia
;    TITLE OF INVENTION:  G-BETA-GAMMA REGULATED
;    TITLE OF INVENTION:  PHOSPHATIDYLINOSITOL-3' KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds, LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036-2811
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/225,170
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/916,917
;      FILING DATE:  15-AUG-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Abrams, Samuel B
;      REGISTRATION NUMBER:  30,605
;      REFERENCE/DOCKET NUMBER:  8549-0006-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-493-4935
;      TELEFAX:  650-493-5556
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5162 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-225-170-13

  Query Match             53.7%;  Score 22;  DB 3;  Length 5162;
  Best Local Similarity   70.0%;  Pred. No. 3.8;
  Matches   28;  Conservative    1;  Mismatches   11;  Indels    0;  Gaps    0;
```

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || || || ||||:  |  |||| ||||||||  |
       5012 CAGGTACTCAGGAGGCTGASGCAGGAGGATTGCCTGAGCC 5051


SULT 10
-09-574-141A-97
Sequence 97, Application US/09574141A
Patent No. 6395490
GENERAL INFORMATION:
 APPLICANT: Gonsalves, Dennis
 APPLICANT:  Meng, Baozhong
 TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
 TITLE OF INVENTION:  NUCLEIC ACIDS, PROTEINS, AND THEIR USES
 FILE REFERENCE: 07678/035005
 CURRENT APPLICATION NUMBER: US/09/574,141A
 CURRENT FILING DATE:  2000-05-18
 PRIOR APPLICATION NUMBER: 60/047,147
 PRIOR FILING DATE: 1997-05-20
 PRIOR APPLICATION NUMBER: 60/069,902
 PRIOR FILING DATE: 1997-12-17
 PRIOR APPLICATION NUMBER: 09/081,320
 PRIOR FILING DATE: 1998-05-19
 NUMBER OF SEQ ID NOS: 97
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 97
  LENGTH: 393
  TYPE: DNA
  ORGANISM: Rupestris stem pitting associated virus
-09-574-141A-97

Query Match           53.2%;  Score 21.8;  DB 4;  Length 393;
Best Local Similarity  78.8%;  Pred. No. 2.3;
Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

          7 GCTGAGAAGCCTGACATACCAGGACTGCCTGAG 39
            |||||| ||  ||||||||   |||||| ||||
        166 GCTGAGCAGAGTGACATACTTAGACTGCTTGAG 198


SULT 11
-09-574-141A-94
Sequence 94, Application US/09574141A
Patent No. 6395490
GENERAL INFORMATION:
 APPLICANT: Gonsalves, Dennis
 APPLICANT:  Meng, Baozhong
 TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
 TITLE OF INVENTION:  NUCLEIC ACIDS, PROTEINS, AND THEIR USES
 FILE REFERENCE: 07678/035005
 CURRENT APPLICATION NUMBER: US/09/574,141A
 CURRENT FILING DATE:  2000-05-18
 PRIOR APPLICATION NUMBER: 60/047,147
 PRIOR FILING DATE: 1997-05-20
 PRIOR APPLICATION NUMBER: 60/069,902
 PRIOR FILING DATE: 1997-12-17
 PRIOR APPLICATION NUMBER: 09/081,320
 PRIOR FILING DATE: 1998-05-19
 NUMBER OF SEQ ID NOS: 97
 SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 94
  LENGTH: 399
  TYPE: DNA
  ORGANISM: Rupestris stem pitting associated virus
-09-574-141A-94

Query Match           53.2%;  Score 21.8;  DB 4;  Length 399;
Best Local Similarity  78.8%;  Pred. No. 2.3;
Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

          7 GCTGAGAAGCCTGACATACCAGGACTGCCTGAG 39
            |||||| ||  ||||||||   |||||| ||||
```

```
Db        169 GCTGAGCAGAGTGACATACTTAGACTGCTTGAG 201


RESULT 12
US-09-081-320-39/c
; Sequence 39, Application US/09081320
; Patent No. 6093544
;  GENERAL INFORMATION:
;    APPLICANT:  Gonsalves, Dennis
;    APPLICANT:  Meng, Baozhong
;    TITLE OF INVENTION:  RUPESTRIS STEM PITTING ASSOCIATED VIRUS
;    TITLE OF INVENTION:  NUCLEIC ACIDS, PROTEINS, AND THEIR USES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/081,320
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/047,147
;      FILING DATE:  20-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/069,902
;      FILING DATE:  17-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1722
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-081-320-39

  Query Match           53.2%;  Score 21.8;  DB 3;  Length 907;
  Best Local Similarity  78.8%;  Pred. No. 2.9;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          7 GCTGAGAAGCCTGACATACCAGGACTGCCTGAG 39
              |||||| ||  ||||||||   |||||| ||||
Db        410 GCTGAGCAGAGTGACATACTTAGACTGCTTGAG 378


RESULT 13
US-09-574-141A-39/c
; Sequence 39, Application US/09574141A
; Patent No. 6395490
; GENERAL INFORMATION:
;  APPLICANT: Gonsalves, Dennis
;  APPLICANT:  Meng, Baozhong
;  TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
;  TITLE OF INVENTION:  NUCLEIC ACIDS, PROTEINS, AND THEIR USES
;  FILE REFERENCE: 07678/035005
;  CURRENT APPLICATION NUMBER: US/09/574,141A
;  CURRENT FILING DATE:  2000-05-18
;  PRIOR APPLICATION NUMBER: 60/047,147
```

PRIOR FILING DATE: 1997-05-20
PRIOR APPLICATION NUMBER: 60/069,902
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 09/081,320
PRIOR FILING DATE: 1998-05-19
NUMBER OF SEQ ID NOS: 97
SOFTWARE: FastSEQ for Windows Version 4.0
EQ ID NO 39
LENGTH: 907
TYPE: DNA
ORGANISM: Rupestris stem pitting associated virus
09-574-141A-39

uery Match 53.2%; Score 21.8; DB 4; Length 907;
est Local Similarity 78.8%; Pred. No. 2.9;
atches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
          7 GCTGAGAAGCCTGACATACCAGGACTGCCTGAG 39
            |||||| ||  ||||||||   |||||| ||||
        410 GCTGAGCAGAGTGACATACTTAGACTGCTTGAG 378
```

ULT 14
09-707-780-39/c
equence 39, Application US/09707780
atent No. 6399308
ENERAL INFORMATION:
APPLICANT: Gonsalves, Dennis
APPLICANT: Meng, Baozhong
TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
FILE REFERENCE: 07678/035006
CURRENT APPLICATION NUMBER: US/09/707,780
CURRENT FILING DATE: 2000-11-07
PRIOR APPLICATION NUMBER: 09/081,320
PRIOR FILING DATE: 1998-05-19
PRIOR APPLICATION NUMBER: 60/047,147
PRIOR FILING DATE: 1997-05-20
PRIOR APPLICATION NUMBER: 60/069,902
PRIOR FILING DATE: 1997-12-17
NUMBER OF SEQ ID NOS: 54
SOFTWARE: FastSEQ for Windows Version 4.0
EQ ID NO 39
LENGTH: 907
TYPE: DNA
ORGANISM: Rupestris stem pitting associated virus
09-707-780-39

uery Match 53.2%; Score 21.8; DB 4; Length 907;
est Local Similarity 78.8%; Pred. No. 2.9;
atches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
          7 GCTGAGAAGCCTGACATACCAGGACTGCCTGAG 39
            |||||| ||  ||||||||   |||||| ||||
        410 GCTGAGCAGAGTGACATACTTAGACTGCTTGAG 378
```

ULT 15
08-965-729A-2/c
equence 2, Application US/08965729A
atent No. 6200751
GENERAL INFORMATION:
APPLICANT: Jian-Ming Gu and Charles T. Esmon
TITLE OF INVENTION: ENDOTHELIUM SPECIFIC EXPRESSION
TITLE OF INVENTION: REGULATED BY EPCR CONTROL ELEMENTS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-4530

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,729A
; FILING DATE: 07-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,718
; FILING DATE: 08-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 164 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3224 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: /note= "Nucleotides 2272 through 2729 are a
; OTHER INFORMATION: thrombin responsive element"; Human
US-08-965-729A-2

Query Match 53.2%; Score 21.8; DB 4; Length 3224;
Best Local Similarity 70.7%; Pred. No. 4.1;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              |||| | || || || ||||  ||  |||| ||| |||| |
Db        198 TCAGCTACTCAGGAGGCTGAGGTAGGAGGATTGCTTGAGCC 158
```

Search completed: July 22, 2003, 11:18:22
Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

nucleic - nucleic search, using sw model

on: July 22, 2003, 11:16:53 ; Search time 130 Seconds
(without alignments)
650.639 Million cell updates/sec

le: U95626-C-AT-42723
fect score: 41
uence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

ring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

rched: 1439767 seqs, 1031500376 residues

al number of hits satisfying chosen parameters: 2879534

mum DB seq length: 0
imum DB seq length: 2000000000

t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	143068	11	US-09-967-768A-316	Sequence 316, App
2	25.2	61.5	591	15	US-10-198-846-11772	Sequence 11772, A
3	24	58.5	544	15	US-10-027-632-66000	Sequence 66000, A
4	24	58.5	544	15	US-10-027-632-66001	Sequence 66001, A
5	24	58.5	544	15	US-10-027-632-66002	Sequence 66002, A
6	24	58.5	544	15	US-10-027-632-298997	Sequence 298997,
7	24	58.5	544	15	US-10-027-632-298998	Sequence 298998,
8	24	58.5	544	15	US-10-027-632-298999	Sequence 298999,
9	24	58.5	552	15	US-10-027-632-286697	Sequence 286697,
10	24	58.5	573	15	US-10-027-632-288310	Sequence 288310,
11	24	58.5	573	15	US-10-027-632-288311	Sequence 288311,
12	24	58.5	17335	11	US-09-764-847-1280	Sequence 1280, Ap
13	24	58.5	17335	15	US-10-092-154-1280	Sequence 1280, Ap
14	24	58.5	19882	11	US-09-764-847-1281	Sequence 1281, Ap
15	24	58.5	19882	15	US-10-092-154-1281	Sequence 1281, Ap
16	24	58.5	183337	15	US-10-020-141-5	Sequence 5, Appli
c 17	24	58.5	202001	10	US-09-734-674-3	Sequence 3, Appli
c 18	24	58.5	202001	15	US-10-274-990-3	Sequence 3, Appli
19	24	58.5	300000	15	US-10-262-552-33	Sequence 33, Appl
20	23.4	57.1	590	15	US-10-027-632-216678	Sequence 216678,
21	23.4	57.1	731	15	US-10-027-632-152011	Sequence 152011,
22	23.4	57.1	2149	15	US-10-027-632-103171	Sequence 103171,
23	23.4	57.1	2149	15	US-10-027-632-103172	Sequence 103172,
c 24	23.4	57.1	38374	11	US-09-880-107-3463	Sequence 3463, Ap
c 25	23	56.1	610	15	US-10-027-632-224761	Sequence 224761,
c 26	22.8	55.6	601	15	US-10-027-632-114311	Sequence 114311,
27	22.6	55.1	413	15	US-10-027-632-55856	Sequence 55856, A
28	22.6	55.1	413	15	US-10-027-632-301953	Sequence 301953,
29	22.6	55.1	676	15	US-10-027-632-222706	Sequence 222706,
30	22.6	55.1	676	15	US-10-027-632-222707	Sequence 222707,
31	22.6	55.1	741	15	US-10-027-632-15461	Sequence 15461, A
32	22.6	55.1	746	15	US-10-027-632-145792	Sequence 145792,
c 33	22.6	55.1	879	15	US-10-027-632-2178	Sequence 2178, Ap
c 34	22.4	54.6	401	15	US-10-027-632-279579	Sequence 279579,
c 35	22.4	54.6	497	15	US-10-027-632-214260	Sequence 214260,
c 36	22.4	54.6	553	15	US-10-027-632-273704	Sequence 273704,
c 37	22.4	54.6	553	15	US-10-027-632-273705	Sequence 273705,
38	22.4	54.6	585	15	US-10-027-632-319580	Sequence 319580,
39	22.4	54.6	585	15	US-10-027-632-319581	Sequence 319581,
40	22.4	54.6	585	15	US-10-027-632-319582	Sequence 319582,
41	22.4	54.6	589	15	US-10-027-632-86169	Sequence 86169, A
c 42	22.4	54.6	598	15	US-10-027-632-206768	Sequence 206768,
43	22.4	54.6	611	15	US-10-027-632-239374	Sequence 239374,
44	22.4	54.6	614	15	US-10-027-632-256903	Sequence 256903,
c 45	22.4	54.6	655	15	US-10-027-632-244369	Sequence 244369,

ALIGNMENTS

RESULT 1
US-09-967-768A-316
; Sequence 316, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 316
; LENGTH: 143068
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-316

Query Match 100.0%; Score 41; DB 11; Length 143068;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||||||||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

RESULT 2
US-10-198-846-11772
; Sequence 11772, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:

APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 11772
LENGTH: 591
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587,
LOCATION: 588, 589, 590, 591
OTHER INFORMATION: n = A,T,C or G
-10-198-846-11772

Query Match 61.5%; Score 25.2; DB 15; Length 591;
Best Local Similarity 78.9%; Pred. No. 0.86;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGA 38
|||| |||| |||||||||||| || || | ||| |||
491 TCAGCTGCTCAGAAGCCTGACGTAGGAGAATTGCTTGA 528

SULT 3
-10-027-632-66000
Sequence 66000, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 66000
LENGTH: 544
TYPE: DNA
ORGANISM: Human
-10-027-632-66000

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||| | || || || |||| | ||||||| ||| ||||||
127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166

RESULT 4
US-10-027-632-66001
; Sequence 66001, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66001
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-66001

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
||| | || || || |||| | ||||||| ||| ||||||
Db 127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166

RESULT 5
US-10-027-632-66002
; Sequence 66002, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66002
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Human

10-027-632-66002

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || || || ||||  | |||||| ||| ||||||
        127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166
```

SULT 6
10-027-632-298997
Sequence 298997, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 298997
LENGTH: 544
TYPE: DNA
ORGANISM: Human
-10-027-632-298997

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || || || ||||  | |||||| ||| ||||||
        127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166
```

SULT 7
-10-027-632-298998
Sequence 298998, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 298998
LENGTH: 544
TYPE: DNA
ORGANISM: Human
US-10-027-632-298998

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | || || || ||||  | |||||| ||| ||||||
Db        127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166
```

RESULT 8
US-10-027-632-298999
; Sequence 298999, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 298999
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-298999

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | || || || ||||  | |||||| ||| ||||||
Db        127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166
```

RESULT 9
US-10-027-632-286697/c
; Sequence 286697, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 286697
LENGTH: 552
TYPE: DNA
ORGANISM: Human
-10-027-632-286697

Query Match 58.5%; Score 24; DB 15; Length 552;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| ||||||| |||| ||| |||| |
        300 CAGCTACTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 261
```

RESULT 10
-10-027-632-288310
Sequence 288310, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 288310
LENGTH: 573
TYPE: DNA
ORGANISM: Human
-10-027-632-288310

Query Match 58.5%; Score 24; DB 15; Length 573;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| ||||||| |||| ||| |||| |
        270 CAGCTACTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 309
```

RESULT 11
-10-027-632-288311
Sequence 288311, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288311
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288311

Query Match 58.5%; Score 24; DB 15; Length 573;
Best Local Similarity 75.0%; Pred. No. 2.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| ||||||| |||| ||| |||| |
Db        270 CAGCTACTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 309
```

RESULT 12
US-09-764-847-1280/c
; Sequence 1280, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1280
; LENGTH: 17335
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1280

Query Match 58.5%; Score 24; DB 11; Length 17335;
Best Local Similarity 75.0%; Pred. No. 3.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | || ||||| |||| | |  |||||||| |||| |
Db       5266 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5227
```

RESULT 13
US-10-092-154-1280/c
; Sequence 1280, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC009C1
CURRENT APPLICATION NUMBER: US/10/092,154
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2003
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1280
LENGTH: 17335
TYPE: DNA
ORGANISM: Homo sapiens
10-092-154-1280

Query Match 58.5%; Score 24; DB 15; Length 17335;
Best Local Similarity 75.0%; Pred. No. 3.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || ||||| |||| | |  |||||||| |||| |
       5266 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5227
```

RESULT 14
09-764-847-1281/c
Sequence 1281, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 2003
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1281
LENGTH: 19882
TYPE: DNA
ORGANISM: Homo sapiens
09-764-847-1281

Query Match 58.5%; Score 24; DB 11; Length 19882;
Best Local Similarity 75.0%; Pred. No. 3.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || ||||| |||| | |  |||||||| |||| |
       5270 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5231
```

RESULT 15
10-092-154-1281/c
Sequence 1281, Application US/10092154
Publication No. US20030054375A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009C1
CURRENT APPLICATION NUMBER: US/10/092,154
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2003
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1281
LENGTH: 19882
TYPE: DNA
ORGANISM: Homo sapiens
10-092-154-1281

Query Match 58.5%; Score 24; DB 15; Length 19882;
Best Local Similarity 75.0%; Pred. No. 3.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | || ||||| |||| | |  |||||||| |||| |
Db        5270 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5231
```

Search completed: July 22, 2003, 12:06:47
Job time : 132 secs

GenCore version 5.1.6

nucleic - nucleic search, using sw model

on: July 22, 2003, 11:16:52 ; Search time 1307 Seconds
(without alignments)
508.045 Million cell updates/sec

le: U95626-C-AT-42723
fect score: 41
quence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 16154066 seqs, 8097743376 residues

tal number of hits satisfying chosen parameters: 32308132

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description
1	27.2	66.3	471	14	BQ300489	BQ300489 PM0-KT004
2	27.2	66.3	538	10	AW653656	AW653656 102451 MA
3	25.6	62.4	315	12	BF935457	BF935457 CM1-NT027
4	25.6	62.4	337	9	AA369692	AA369692 EST81420
5	25.6	62.4	384	12	BF990363	BF990363 CM0-GN016
6	25.6	62.4	384	12	BF995389	BF995389 CM0-GN016
7	25.6	62.4	384	12	BG002012	BG002012 CM0-GN016
8	25.6	62.4	423	12	BF760350	BF760350 CM1-CT060
9	25.6	62.4	470	12	BG002336	BG002336 QV4-GN025
10	25.6	62.4	493	14	BQ300457	BQ300457 PM0-KT004
c 11	25.6	62.4	509	17	AQ898508	AQ898508 HS_3127_B
12	25.6	62.4	512	12	BF858648	BF858648 RC5-FT019
c 13	25.6	62.4	841	14	BQ719041	BQ719041 AGENCOURT
c 14	25.2	61.5	440	9	AI140599	AI140599 qe05e09.x
15	25.2	61.5	529	12	BF891477	BF891477 PM3-MT011
c 16	25.2	61.5	753	10	AW978551	AW978551 EST390660
17	25	61.0	352	17	B85322	B85322 RPCI11-16C1
18	25	61.0	368	17	AQ132663	AQ132663 HS_3051_B
19	24	58.5	214	10	BB596733	BB596733 BB596733
c 20	24	58.5	277	12	BG007496	BG007496 RC2-GN025
21	24	58.5	280	9	AA372587	AA372587 EST84770
c 22	24	58.5	309	12	BF931001	BF931001 IL2-NT020
23	24	58.5	322	12	BF926436	BF926436 RC6-NT015
c 24	24	58.5	361	14	BQ345670	BQ345670 PM3-NT031
25	24	58.5	374	10	AW386852	AW386852 CM0-PT004
26	24	58.5	380	13	BI025579	BI025579 CM3-MT029
27	24	58.5	399	10	AV730530	AV730530 AV730530
c 28	24	58.5	412	17	AQ181316	AQ181316 HS_3223_A
29	24	58.5	424	10	AV729997	AV729997 AV729997
30	24	58.5	457	17	AQ423186	AQ423186 CITBI-E1-
c 31	24	58.5	470	12	BF904944	BF904944 IL2-MT017
32	24	58.5	480	12	BF997233	BF997233 PM0-GN019
33	24	58.5	509	12	BE933206	BE933206 RC1-HT088
34	24	58.5	524	17	AQ784899	AQ784899 HS_3251_A
c 35	24	58.5	543	14	BQ323349	BQ323349 QV0-CI001
c 36	24	58.5	560	17	AQ419838	AQ419838 RPCI-11-1
c 37	24	58.5	561	13	BG986513	BG986513 MR1-HT118
c 38	24	58.5	571	14	BQ323273	BQ323273 QV0-CI001
39	24	58.5	571	17	AQ504035	AQ504035 RPCI-11-2
40	24	58.5	586	9	AL708735	AL708735 DKFZp686P
c 41	24	58.5	592	12	BF819216	BF819216 QV2-CI010
c 42	24	58.5	602	10	BE144468	BE144468 MR0-HT016
43	24	58.5	636	17	AQ616424	AQ616424 HS_5149_A
44	24	58.5	648	17	AG147167	AG147167 Pan trogl
45	24	58.5	663	17	AG149846	AG149846 Pan trogl

ALIGNMENTS

RESULT 1
BQ300489
LOCUS BQ300489 471 bp mRNA linear EST 16-MAY-2002
DEFINITION PM0-KT0041-070201-004-b04 KT0041 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ300489
VERSION BQ300489.1 GI:20816011
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 471)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

```
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-KT0041-
          070201-004-b04&t3=2001-02-07&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 11
          High quality sequence stop: 468.
ATURES              Location/Qualifiers
   source           1. .471
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="KT0041"
                    /dev_stage="Adult"
                    /note="Organ: bladder tumor; Vector: puc18; Site_1: SmaI;
                    Site_2: SmaI; A mini-library was made by cloning products
                    derived from  ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 -  Ludwig Institute for Cancer Research)
                    profiles into  the pUC 18 vector. Reverse transcription of
                    tissue mRNA and cDNA amplification were performed under
                    low stringency conditions."
SE COUNT      133 a    119 c    115 g    104 t
IGIN

Query Match           66.3%;  Score 27.2;  DB 14;  Length 471;
Best Local Similarity   80.0%;  Pred. No. 7.2;
Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||||||||| |||||||   |||||||| |||| |
        148 CAGCTACTGAGAAGGCTGACATGGTAGGACTGCTTGAGCC 187


SULT 2
653656
CUS         AW653656                538 bp    mRNA    linear   EST 25-APR-2001
FINITION    102451 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
CESSION     AW653656
RSION       AW653656.1  GI:7419482
YWORDS      EST.
URCE        cow.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
FERENCE     1  (bases 1 to 538)
AUTHORS     Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
TITLE       Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL     Genome Res. 11 (4), 626-630 (2001)
MEDLINE     21180013
MMENT       Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 107  row: K  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
ATURES              Location/Qualifiers
   source           1. .538
                    /organism="Bos taurus"
                    /db_xref="taxon:9913"
                    /clone_lib="MARC 1BOV"
                    /tissue_type="pooled"
                    /lab_host="DH10B"
                    /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                    Library made from pooled tissue from lymph node, ovary,
                    fat, hypothalamus, and pituitary."
BASE COUNT      150 a    132 c    123 g    133 t
ORIGIN

  Query Match           66.3%;  Score 27.2;  DB 10;  Length 538;
  Best Local Similarity   80.0%;  Pred. No. 7.5;
  Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGA 40
              |||||||||||||||||||| |||||  | | | | |
Db        362 TCAGTTGCTGAGAAGCCTGACACACCAGTGCCACTTAAAA 401


RESULT 3
BF935457
LOCUS       BF935457                315 bp    mRNA    linear   EST 22-JAN-2001
DEFINITION  CM1-NT0270-271200-678-e04 NT0270 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF935457
VERSION     BF935457.1  GI:12352768
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM1&t2=CM1-NT0270-
            271200-678-e04&t3=2000-12-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 15
            High quality sequence stop: 315.
FEATURES            Location/Qualifiers
     source         1. .315
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="NT0270"
                    /dev_stage="Adult"
                    /note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
                    Site_2: SmaI; A mini-library was made by cloning products
                    derived from  ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 -  Ludwig Institute for Cancer Research)
                    profiles into  the pUC 18 vector. Reverse transcription of
                    tissue mRNA and cDNA amplification were performed under
                    low stringency conditions."
BASE COUNT       78 a     70 c     88 g     79 t
ORIGIN

  Query Match           62.4%;  Score 25.6;  DB 12;  Length 315;
  Best Local Similarity   77.5%;  Pred. No. 24;
```

```
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| |||||||   |||||||| |||| |
        246 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 285


SULT 4
369692
CUS       AA369692                337 bp    mRNA    linear   EST 21-APR-1997
FINITION  EST81420 Prostate gland I Homo sapiens cDNA 3' end similar to EST
          containing Alu repeat, mRNA sequence.
CESSION   AA369692
RSION     AA369692.1  GI:2022225
YWORDS    EST.
URCE      human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 337)
AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
ITLE      Rapid cDNA sequencing (expressed sequence tags) from a
          directionally cloned human infant brain cDNA library
JOURNAL   Nat. Genet. 4, 373-380 (1993)
MEDLINE   94004965
MMENT     Other_ESTs: EST81421 THC100858
          Contact: Kerlavage, AR
          Bioinformatics
          The Institute for Genomic Research
          9712 Medical Center Drive, Rockville, MD 20850 USA
          Tel: 3018699056
          Fax: 3018699423
          Email: arkerlav@tigr.org
          For clone availability, additional sequence and expression
          information related to this EST, please check the TIGR Human Gene
          Index (http://www.tigr.org/tdb/hgi/hgi.html)
          Seq primer: M13-21.
ATURES              Location/Qualifiers
   source           1. .337
                    /organism="Homo sapiens"
                    /db_xref="ATCC (inhost):174463"
                    /db_xref="taxon:9606"
                    /clone_lib="Prostate gland I"
                    /sex="male"
                    /dev_stage="adult, 21 yrs"
                    /note="Organ: prostate; Vector: pBluescript SK-; Site_1:
                    EcoRI; Site_2: XhoI"
SE COUNT       76 a     79 c     97 g     84 t      1 others
IGIN

Query Match             62.4%;  Score 25.6;  DB 9;  Length 337;
Best Local Similarity   77.5%;  Pred. No. 25;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| |||||||   |||||||| |||| |
         94 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 133


SULT 5
990363
CUS       BF990363                384 bp    mRNA    linear   EST 23-JAN-2001
FINITION  CM0-GN0163-251000-623-a02 GN0163 Homo sapiens cDNA, mRNA sequence.
CESSION   BF990363
RSION     BF990363.1  GI:12396688
YWORDS    EST.
URCE      human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 384)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-GN0163-
          251000-623-a02&t3=2000-10-25&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 383.
FEATURES            Location/Qualifiers
     source         1. .384
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="GN0163"
                    /dev_stage="Adult"
                    /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                    ; Site_2: SmaI; A mini-library was made by cloning
                    products derived from  ORESTES PCR (U.S. Letters Patent
                    application No. 196,716 -  Ludwig Institute for Cancer
                    Research) profiles into  the pUC 18 vector. Reverse
                    transcription of tissue mRNA and cDNA amplification were
                    performed under low stringency conditions."
BASE COUNT       88 a     85 c    104 g    107 t
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 12;  Length 384;
  Best Local Similarity   77.5%;  Pred. No. 26;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| ||||| |||| |||||||   |||| ||| |||| |
Db        137 CAGCTGCTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 176


RESULT 6
BF995389
LOCUS       BF995389                384 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  CM0-GN0163-041100-623-a02 GN0163 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF995389
VERSION     BF995389.1  GI:12401712
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
```

```
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-GN0163-
          041100-623-a02&t3=2000-11-04&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 383.
ATURES              Location/Qualifiers
   source           1. .384
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="GN0163"
                    /dev_stage="Adult"
                    /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                    ; Site_2: SmaI; A mini-library was made by cloning
                    products derived from  ORESTES PCR (U.S. Letters Patent
                    application No. 196,716 -  Ludwig Institute for Cancer
                    Research) profiles into  the pUC 18 vector. Reverse
                    transcription of tissue mRNA and cDNA amplification were
                    performed under low stringency conditions."
SE COUNT       88 a     85 c    104 g    107 t
IGIN

Query Match             62.4%;  Score 25.6;  DB 12;  Length 384;
Best Local Similarity   77.5%;  Pred. No. 26;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| ||||| |||| |||||||   |||| ||| |||| |
        137 CAGCTGCTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 176


SULT 7
002012
CUS       BG002012                384 bp    mRNA    linear   EST 24-JAN-2001
FINITION  CM0-GN0162-151100-623-a02 GN0162 Homo sapiens cDNA, mRNA sequence.
CESSION   BG002012
RSION     BG002012.1  GI:12440900
YWORDS    EST.
URCE      human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 384)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
MMENT     Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-GN0162-
          151100-623-a02&t3=2000-11-15&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 383.
ATURES              Location/Qualifiers
   source           1. .384
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="GN0162"
                    /dev_stage="Adult"
                    /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                    ; Site_2: SmaI; A mini-library was made by cloning
                    products derived from  ORESTES PCR (U.S. Letters Patent
                    application No. 196,716 -  Ludwig Institute for Cancer
                    Research) profiles into  the pUC 18 vector. Reverse
                    transcription of tissue mRNA and cDNA amplification were
                    performed under low stringency conditions."
BASE COUNT       88 a     85 c    104 g    107 t
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 12;  Length 384;
  Best Local Similarity   77.5%;  Pred. No. 26;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| ||||| |||| |||||||   |||| ||| |||| |
Db        137 CAGCTGCTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 176


RESULT 8
BF760350
LOCUS       BF760350                423 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  CM1-CT0605-211200-671-b09 CT0605 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF760350
VERSION     BF760350.1  GI:12108250
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM1&t2=CM1-CT0605-
            211200-671-b09&t3=2000-12-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 7
            High quality sequence stop: 423.
FEATURES            Location/Qualifiers
     source         1. .423
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="CT0605"
                    /dev_stage="Adult"
                    /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                    SmaI; A mini-library was made by cloning products derived
                    from  ORESTES PCR (U.S. Letters Patent application No. 196
                    ,716 -  Ludwig Institute for Cancer Research) profiles
                    into  the pUC 18 vector. Reverse transcription of tissue
                    mRNA and cDNA amplification were performed under low
                    stringency conditions."
```

```
E COUNT      119 a     92 c    110 g    102 t
GIN

uery Match             62.4%;  Score 25.6;  DB 12;  Length 423;
est Local Similarity   77.5%;  Pred. No. 27;
atches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| |||||||   |||||||| |||| |
        253 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 292


ULT 9
02336
US       BG002336                 470 bp    mRNA    linear   EST 24-JAN-2001
INITION  QV4-GN0250-161100-543-c01 GN0250 Homo sapiens cDNA, mRNA sequence.
ESSION   BG002336
SION     BG002336.1  GI:12441546
WORDS    EST.
RCE      human.
RGANISM  Homo sapiens
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ERENCE   1  (bases 1 to 470)
UTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
         Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
         Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
         Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
         ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
         Simpson,A.J.
ITLE     Shotgun sequencing of the human transcriptome with ORF expressed
         sequence tags
OURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
EDLINE   20202663
MENT     Contact: Simpson A.J.G.
         Laboratory of Cancer Genetics
         Ludwig Institute for Cancer Research
         Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
         Brazil
         Tel: +55-11-2704922
         Fax: +55-11-2707001
         Email: asimpson@ludwig.org.br
         This sequence was derived from the FAPESP/LICR Human Cancer Genome
         Project. This entry can be seen in the following URL
         (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV4&t2=QV4-GN0250-
         161100-543-c01&t3=2000-11-16&t4=1)
         Seq primer: puc 18 forward
         High quality sequence start: 2
         High quality sequence stop: 470.
TURES             Location/Qualifiers
  source          1. .470
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_lib="GN0250"
                  /dev_stage="Adult"
                  /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                  ; Site_2: SmaI; A mini-library was made by cloning
                  products derived from  ORESTES PCR (U.S. Letters Patent
                  application No. 196,716 -  Ludwig Institute for Cancer
                  Research) profiles into  the pUC 18 vector. Reverse
                  transcription of tissue mRNA and cDNA amplification were
                  performed under low stringency conditions."
E COUNT      119 a    104 c    125 g    122 t
GIN

uery Match             62.4%;  Score 25.6;  DB 12;  Length 470;
est Local Similarity   77.5%;  Pred. No. 28;
atches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||| | ||| |||| |||||||   |||||||| |||| |
         62 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 101
```

```
RESULT 10
BQ300457
LOCUS       BQ300457                 493 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  PM0-KT0041-030201-003-h02 KT0041 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ300457
VERSION     BQ300457.1  GI:20815979
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-KT0041-
            030201-003-h02&t3=2001-02-03&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 14
            High quality sequence stop: 467.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="KT0041"
                     /dev_stage="Adult"
                     /note="Organ: bladder_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      137 a    126 c    120 g    110 t
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 14;  Length 493;
  Best Local Similarity   77.5%;  Pred. No. 28;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   |||||||| |||| |
Db        187 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 226


RESULT 11
AQ898508/c
LOCUS       AQ898508                 509 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_3127_B2_C10_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3127 Col=20 Row=F, DNA sequence.
ACCESSION   AQ898508
VERSION     AQ898508.1  GI:6354698
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ERENCE    1  (bases 1 to 509)
UTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
          Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
          Hood,L.
ITLE      Sequence-tagged connectors: A sequence approach to mapping and
          scanning the human genome
OURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
EDLINE    99380589
MENT      Contact: Mahairas GG, Wallace JC, Hood L
          High Throughput Sequencing Center
          University of Washington
          401 Queen Anne Avenue North, Seattle, WA 98109, USA
          Tel: (206) 616-3618
          Fax: (206) 616-3887
          Email: jwallace@u.washington.edu
          Clones may be purchased from Research Genetics (info@resgen.com).
          BAC end Web Server: http://www.htsc.washington.edu
          Plate: 3127  row: F  column: 20
          Seq primer: T7
          Class: BAC ends
          High quality sequence stop: 509.
TURES               Location/Qualifiers
   source           1. .509
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Plate=3127 Col=20 Row=F"
                    /clone_lib="CIT Approved Human Genomic Sperm Library D"
                    /sex="male"
                    /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                    E-Coli DH10B"
SE COUNT      125 a    121 c    103 g    154 t      6 others
IGIN

Query Match             62.4%;  Score 25.6;  DB 17;  Length 509;
Best Local Similarity   77.5%;  Pred. No. 29;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
            ||||| || ||||| ||||  || ||||| ||| |||| |
        237 CAGTTACTCAGAAGGCTGAGGTAGCAGGATTGCTTGAGCC 198


ULT 12
858648
CUS       BF858648                512 bp    mRNA    linear   EST 16-JAN-2001
INITION   RC5-FT0194-271100-022-A06 FT0194 Homo sapiens cDNA, mRNA sequence.
CESSION   BF858648
RSION     BF858648.1  GI:12246392
YWORDS    EST.
URCE      human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 512)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
ITLE      Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
OURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
EDLINE    20202663
MENT      Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC5&t2=RC5-FT0194-
          271100-022-A06&t3=2000-11-27&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 73
          High quality sequence stop: 506.
FEATURES            Location/Qualifiers
     source         1. .512
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="FT0194"
                    /dev_stage="Adult"
                    /note="Organ: prostate tumor; Vector: puc18; Site_1: SmaI;
                    Site_2: SmaI; A mini-library was made by cloning products
                    derived from  ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 -  Ludwig Institute for Cancer Research)
                    profiles into  the pUC 18 vector. Reverse transcription of
                    tissue mRNA and cDNA amplification were performed under
                    low stringency conditions."
BASE COUNT      146 a    110 c    138 g    118 t
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 12;  Length 512;
  Best Local Similarity   77.5%;  Pred. No. 29;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   |||||||| |||| |
Db        257 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 296


RESULT 13
BQ719041/c
LOCUS       BQ719041                841 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8100912 Lupski_sympathetic_trunk Homo sapiens cDNA clone
            IMAGE:6189820 5', mRNA sequence.
ACCESSION   BQ719041
VERSION     BQ719041.1  GI:21857938
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13587  row: k  column: 05
            High quality sequence stop: 615.
FEATURES            Location/Qualifiers
     source         1. .841
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:6189820"
                    /clone_lib="Lupski_sympathetic_trunk"
                    /sex="male"
                    /tissue_type="sympathetic trunk"
                    /dev_stage="adult, 16 yr"
                    /lab_host="DH10B"
                    /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
                    NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                    Directionally cloned using the following adaptors:
```

5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

BASE COUNT 187 a 221 c 207 g 226 t
ORIGIN

Query Match 62.4%; Score 25.6; DB 14; Length 841;
Best Local Similarity 77.5%; Pred. No. 34;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||||||   |||||||| |||| |
Db        361 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 322
```

RESULT 14
AI140599/c
LOCUS AI140599 440 bp mRNA linear EST 29-OCT-1998
DEFINITION qe05e09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1738120 3' similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION AI140599
VERSION AI140599.1 GI:3648056
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 440)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 377 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 427.
FEATURES Location/Qualifiers
source 1. .440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1738120"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 103 a 115 c 60 g 161 t 1 others
ORIGIN

Query Match 61.5%; Score 25.2; DB 9; Length 440;
Best Local Similarity 78.9%; Pred. No. 38;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGA 38
              |||| |||| |||||||||||  ||  || | ||| |||
Db        101 TCAGCTGCTCAGAAGCCTGACGTAGGAGAATTGCTTGA 64
```

RESULT 15
BF891477
LOCUS BF891477 529 bp mRNA linear EST 18-JAN-2001
DEFINITION PM3-MT0110-241000-010-g07 MT0110 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF891477
VERSION BF891477.1 GI:12282936
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-MT0110-241000-010-g07&t3=2000-10-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 49
High quality sequence stop: 529.
FEATURES Location/Qualifiers
source 1. .529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0110"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 161 a 102 c 110 g 156 t
ORIGIN

Query Match 61.5%; Score 25.2; DB 12; Length 529;
Best Local Similarity 78.9%; Pred. No. 41;
Matches 30; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGA 38
              |||| |||| |||||||||||  ||  || | ||| |||
Db        152 TCAGCTGCTCAGAAGCCTGACGTAGGAGAATTGCTTGA 189
```

Search completed: July 22, 2003, 12:02:13
Job time : 1311 secs

GenCore version 5.1.6

ucleic - nucleic search, using sw model

on: July 22, 2003, 11:16:54 ; Search time 985 Seconds
(without alignments)
1211.386 Million cell updates/sec

e: U95626-A-AT-42723
fect score: 41
ence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

ing table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ched: 2054640 seqs, 14551402878 residues

l number of hits satisfying chosen parameters: 4109280

mum DB seq length: 0
mum DB seq length: 2000000000

-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39.4	96.1	583	9	HSMCPA01	U80923 Human monoc
2	39.4	96.1	2082	9	AF068265	AF068265 Homo sapi
3	39.4	96.1	143068	6	AX335952	AX335952 Sequence
4	39.4	96.1	143068	9	HSU95626	U95626 Homo sapien
5	39.4	96.1	185437	9	AC098613	AC098613 Homo sapi
6	27.2	66.3	120793	2	AF322450	AF322450 Homo sapi
7	27.2	66.3	133544	9	AC100748	AC100748 Homo sapi
8	27.2	66.3	163035	2	AC006405	AC006405 Homo sapi
9	26.6	64.9	82601	9	AL136971	AL136971 Human DNA
c 10	26.6	64.9	142441	2	HSJ344H20	AL132710 Homo sapi
11	26.6	64.9	202750	2	AC116942	AC116942 Pan trogl
c 12	25.6	62.4	59155	2	AC104983	AC104983 Homo sapi
13	25.6	62.4	110394	9	AC003086	AC003086 Homo sapi
14	25.6	62.4	111312	9	AC011248	AC011248 Homo sapi
15	25.6	62.4	142082	2	AC093137	AC093137 Pan trogl
c 16	25.6	62.4	159365	9	AL392111	AL392111 Human DNA
c 17	25.6	62.4	171267	2	AC040979	AC040979 Homo sapi
c 18	25.6	62.4	176738	9	AC090960	AC090960 Homo sapi
19	25.6	62.4	178451	9	AL139396	AL139396 Human DNA
c 20	25.6	62.4	178942	9	AL357520	AL357520 Human DNA
21	25.6	62.4	181755	2	AC013283	AC013283 Homo sapi
22	25.6	62.4	190948	2	AC093138	AC093138 Pan trogl
23	25.6	62.4	208392	2	AC084876	AC084876 Homo sapi
c 24	25.6	62.4	210736	2	AC092792	AC092792 Homo sapi
25	25.6	62.4	214304	2	AC026112	AC026112 Homo sapi
26	25.2	61.5	132463	2	AC111775	AC111775 Rattus no
27	25.2	61.5	187803	10	AL596252	AL596252 Mouse DNA
28	25	61.0	41615	9	AC005625	AC005625 Homo sapi
29	25	61.0	68116	2	AC103970	AC103970 Homo sapi
c 30	25	61.0	73235	2	AC103725	AC103725 Homo sapi
31	25	61.0	78986	2	AC115626	AC115626 Homo sapi
c 32	25	61.0	112204	9	AL136439	AL136439 Human DNA
33	25	61.0	113836	9	AL133174	AL133174 Human DNA
c 34	25	61.0	116845	9	AC008819	AC008819 Homo sapi
35	25	61.0	117583	9	AC034238	AC034238 Homo sapi
36	25	61.0	120689	9	CNS01RGS	AL159191 Human chr
c 37	25	61.0	133322	2	AC091845	AC091845 Homo sapi
38	25	61.0	143506	9	AC092833	AC092833 Homo sapi
c 39	25	61.0	145871	9	AL391601	AL391601 Human DNA
c 40	25	61.0	151122	2	AF239614	AF239614 Homo sapi
c 41	25	61.0	152141	2	AC090189	AC090189 Homo sapi
c 42	25	61.0	155428	9	AC009682	AC009682 Homo sapi
c 43	25	61.0	159185	2	AC013550	AC013550 Homo sapi
c 44	25	61.0	160979	2	AC055879	AC055879 Homo sapi
45	25	61.0	164125	9	AC018693	AC018693 Homo sapi

ALIGNMENTS

RESULT 1
HSMCPA01
LOCUS HSMCPA01 583 bp DNA linear PRI 24-JUL-1997
DEFINITION Human monocyte chemoattractant protein 1 receptor gene, 5' region.
ACCESSION U80923
VERSION U80923.1 GI:1773031
KEYWORDS .
SEGMENT 1 of 2
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Wong,L.M., Myers,S.J., Tsou,C.L., Gosling,J., Arai,H. and
Charo,I.F.

```
TITLE     Organization and differential expression of the human monocyte
          chemoattractant protein 1 receptor gene. Evidence for the role of
          the carboxyl-terminal tail in receptor trafficking
JOURNAL   J. Biol. Chem. 272 (2), 1038-1045 (1997)
MEDLINE   97150864
PUBMED    8995400
REFERENCE 2  (bases 1 to 583)
AUTHORS   Myers,S.J. and Charo,I.F.
TITLE     Direct Submission
JOURNAL   Submitted (04-DEC-1996) Pharmacology, Emory University, 1510
          Clifton Road, Atlanta 30322, USA
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      157 a    136 c    156 g    134 t
ORIGIN

Query Match             96.1%;  Score 39.4;  DB 9;  Length 583;
Best Local Similarity   97.6%;  Pred. No. 5.2e-05;
Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||| ||||||||||||||||||||
Db         62 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 102


RESULT 2
AF068265
LOCUS       AF068265               2082 bp    DNA     linear   PRI 20-APR-1999
DEFINITION  Homo sapiens monocyte chemoattractant protein 1 receptor (CCR2)
          gene promoter and mRNA, partial sequence.
ACCESSION   AF068265
VERSION     AF068265.1  GI:4587865
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2082)
  AUTHORS   Yamamoto,K., Takeshima,H., Hamada,K., Nakao,M., Kino,T., Nishi,T.,
          Kochi,M., Kuratsu,J., Yoshimura,T. and Ushio,Y.
  TITLE     Cloning and functional characterization of the 5'-flanking region
          of the human monocyte chemoattractant protein-1 receptor (CCR2)
          gene. Essential role of 5'-untranslated region in tissue-specific
          expression
  JOURNAL   J. Biol. Chem. 274 (8), 4646-4654 (1999)
  MEDLINE   99143121
  PUBMED    9988701
REFERENCE   2  (bases 1 to 2082)
  AUTHORS   Yamamoto,K., Takeshima,H., Hamada,K., Nakao,M., Kino,T., Nishi,T.,
          Kochi,M., Kuratsu,J.-I., Yoshimura,T. and Ushio,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) Neurosurgery, Kumamoto University School of
          Medicine, Honjo, 1-1-1, Kumamoto 860-8556, Japan
FEATURES             Location/Qualifiers
     source          1. .2082
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p"
     gene            1. .>2082
                     /gene="CCR2"
     promoter        1. .1697
                     /gene="CCR2"
     CAAT_signal     1570. .1574
                     /gene="CCR2"
     TATA_signal     1666. .1672
                     /gene="CCR2"
     mRNA            1698. .>2012
                     /gene="CCR2"
                     /product="monocyte chemoattractant protein 1 receptor"
     5'UTR           1698. .2012
                     /gene="CCR2"
BASE COUNT      650 a    439 c    433 g    560 t
ORIGIN

Query Match             96.1%;  Score 39.4;  DB 9;  Length 2082;
Best Local Similarity   97.6%;  Pred. No. 4.6e-05;
Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||| ||||||||||||||||||||
Db       1959 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 1999


RESULT 3
AX335952
LOCUS       AX335952               143068 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 6461 from Patent WO0194629.
ACCESSION   AX335952
VERSION     AX335952.1  GI:18126671
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
          Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
          gene sets
  JOURNAL   Patent: WO 0194629-A 6461 13-DEC-2001;
          Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .143068
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    41194 a  30122 c  32403 g  39349 t
ORIGIN

Query Match             96.1%;  Score 39.4;  DB 6;  Length 143068;
Best Local Similarity   97.6%;  Pred. No. 3e-05;
Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||| ||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743


RESULT 4
HSU95626
LOCUS       HSU95626               143068 bp    DNA     linear   PRI 16-MAY-1997
DEFINITION  Homo sapiens ccr2b (ccr2), ccr2a (ccr2), ccr5 (ccr5) and ccr6
          (ccr6) genes, complete cds, and lactoferrin (lactoferrin) gene,
          partial cds, complete sequence.
ACCESSION   U95626
VERSION     U95626.1  GI:2104517
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143068)
  AUTHORS   McCombie,W.R., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D.,
          Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K.,
          Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J.,
          Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M.,
          Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E.,
          Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y.,
          Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.
  TITLE     Human BAC clone 110P12
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 143068)
  AUTHORS   McCombie,R.W., Wilson,R., Chen,E., Gibbs,R., Zuo,L., Johnson,D.,
          Nhan,M., Parnell,L., Dedhia,N., Ansari,A., Mardis,E., Schutz,K.,
```

Gnoj,L., de la Bastide,M., Kaplan,N., Greco,T., Touchman,J., Muzny,D., Chen,C.-N., Evans,C., FitzGerald,M., See,L.H., Tang,M., Porcel,B.M., Dragan,Y., Giacalone,J., Pae,A., Powell,E., Solinsky,K.A., DeSilva,U., Diaz-Perez,S., Zhou,X., Yu,Y., Watanabe,M., Doggett,N., Garcia,D. and Sagripanti,J.-L.

'ITLE Direct Submission

IOURNAL Submitted (27-MAR-1997) Advanced Genome Sequence Analysis Course, Cold Spring Harbor Laboratory, 1Bungtown Rd., Cold Spring Harbor, NY 11724, USA

IMENT Regions with single-strand coverage are as follows:

31434 - 31443 37900 - 37968 53303 - 53357
59166 - 59206 63708 - 63998 65200 - 65335
78605 - 78713 92135 - 92137 112377 - 112551
112643 - 112778 134284 - 134309 134914 - 135019
143046 - 144068.

\TURES Location/Qualifiers

source 1. .143068
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="BAC 110P12"

gene 46056. .49505
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) mRNA (Accession Number U80924), two alternatively spliced mRNAs."

mRNA join(46056. .47046,48255. .49505)
/gene="ccr2"
/product="ccr2a"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced mRNA encoding A-form carboxyl tail, Accession Number U80924."

mRNA 46056. .47997
/gene="ccr2"
/product="ccr2b"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced mRNA encoding B-form carboxyl tail. Accession Number: U80924."

CDS join(46106. .47046,48255. .48438)
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2) alternatively spliced A-form, Encoded by GenBank Accession Number U80924, gi 1168965"
/codon_start=1
/product="ccr2a"
/protein_id="AAB57791.1"
/db_xref="GI:2104518"
/translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRSLFHIALGCRIAPLQKPVCGG
PGVRPGKNVKVTTQGLLDGRGKGKSIGRAPEASLQDKEGA"

CDS 46106. .47188
/gene="ccr2"
/note="confirmed by similarity to Human monocyte chemoattractant protein 1 receptor (ccr2), Accession Number: 1168965"
/codon_start=1
/product="ccr2b"
/protein_id="AAB57792.1"
/db_xref="GI:2104519"
/translation="MLSTSRSRFIRNTNESGEEVTTFFDYDYGAPCHKFDVKQIGAQL
LPPLYSLVFIFGFVGNMLVVLILINCKKLKCLTDIYLLNLAISDLLFLITLPLWAHSA
ANEWVFGNAMCKLFTGLYHIGYFGGIFFIILLTIDRYLAIVHAVFALKARTVTFGVVT
SVITWLVAVFASVPGIIFTKCQKEDSVYVCGPYFPRGWNNFHTIMRNILGLVLPLLIM
VICYSGILKTLLRCRNEKKRHRAVRVIFTIMIVYFLFWTPYNIVILLNTFQEFFGLSN
CESTSQLDQATQVTETLGMTHCCINPIIYAFVGEKFRRYLSVFFRKHITKRFCKQCPV
FYRETVDGVTSTNTPSTGEQEVSAGL"

gene 59531. .64785
/gene="ccr5"

mRNA join(59531. .59573,61472. .64785)
/gene="ccr5"
/product="ccr5"
/note="confirmed by similarity to Human cc chemokine receptor 5 (ccr5) mRNA. Accession number: U54994."

CDS 61483. .62541
/gene="ccr5"
/note="confirmed by similarity to human CC chemokine receptor 5 (ccr5) protein, encoded by GenBank Accession Number U54994, gi 1457946"
/codon_start=1
/product="ccr5"
/protein_id="AAB57793.1"
/db_xref="GI:2104520"
/translation="MDYQVSSPIYDINYYTSEPCQKINVKQIAARLLPPLYSLVFIFG
FVGNMLVILILINCKRLKSMTDIYLLNLAISDLFFLLTVPFWAHYAAAQWDFGNTMCQ
LLTGLYFIGFFSGIFFIILLTIDRYLAVVHAVFALKARTVTFGVVTSVITWVVAVFAS
LPGIIFTRSQKEGLHYTCSSHFPYSQYQFWKNFQTLKIVILGLVLPLLVMVICYSGIL
KTLLRCRNEKKRHRAVRLIFTIMIVYFLFWAPYNIVLLLNTFQEFFGLNNCSSSNRLD
QAMQVTETLGMTHCCINPIIYAFVGEKFRNYLLVFFQKHIAKRFCKCCSIFQQEAPER
ASSVYTRSTGEQEISVGL"

mRNA 96634. .97683
/product="ccr6"
/note="Putative mRNA identified by homology to CCR5 mRNA."
/evidence=not_experimental

gene 96642. .97676
/gene="ccr6"

CDS 96642. .97676
/gene="ccr6"
/note="Identified as a gene by Grail Version 1.3c. Translated sequence exhibits 42% sequence identity to CCR5 protein."
/codon_start=1
/evidence=not_experimental
/product="ccr6"
/protein_id="AAB57794.1"
/db_xref="GI:2104521"
/translation="MANYTLAPEDEYDVLIEGELESDEAEQCDKYDAQALSAQLVPSL
CSAVFVIGVLDNLLVVLILVKYKGLKRVENIYLLNLAVSNLCFLLTLPFWAHAGGDPM
CKILIGLYFVGLYSETFFNCLLTVQRYLVFLHKGNFFSARRRVPCGIITSVLAWVTAI
LATLPEYVVYKPQMEDQKYKCAFSRTPFLPADETFWKHFLTLKMNISVLVLPLFIFTF
LYVQMRKTLRFREQRYSLFKLVFAIMVVFLLMWAPYNIAFFLSTFKEHFSLSDCKSSY
NLDKSVHITKLIATTHCCINPLLYAFLDGTFSKYLCRCFHLRSNTPLQPRGQSAQGTS
REEPDHSTEV"

gene complement(124605. .>139255)
/gene="lactoferrin"

mRNA complement(join(124605. .124816,126528. .126717, 127884. .128068,130006. .130073,132023. .132164, 133863. .134018,135022. .135075,135890. .135980, 137445. .137599,138436. .138610,139077. .>139255))
/gene="lactoferrin"
/product="lactoferrin"
/note="confirmed by similarity to lactoferrin mRNA, accession number M73700"

CDS complement(join(124782. .124816,126528. .126717, 127884. .128068,130006. .130073,132023. .132164, 133863. .134018,135022. .135075,135890. .135980, 137445. .137599,138436. .138610,139077. .>139253))
/gene="lactoferrin"
/note="confirmed by similarity to lactoferrin protein, encoded by GenBank Accession Number M73700, gi 186818"
/codon_start=1
/product="lactoferrin"
/protein_id="AAB57795.1"
/db_xref="GI:2104522"
/translation="DLSDEAERDEYELLCPDNTRKPVDKFKDCHLARVPSHAVVARSV
NGKEDAIWNLLRQAQEKFGKDKSPKFQLFGSPSGQKDLLFKDSAIGFSRVPPRIDSGL
YLGSGYFTAIQNLRKSEEEVAARRARVVWCAVGEQELRKCNQWSGLSEGSVTCSSAST
TEDCIALVLKGEADAMSLDGGYVYTAGKCGLVPVLAENYKSQQSSDPDPNCVDRPVEG
YLAVAVVRRSDTSLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQTGSCKFDEYFSQS

```
                CAPGSDPRSNLCALCIGDEQGENKCVPNSNERYYGYTGAFRCLAENAGDVAFVKDVTV
                LQNTDGNNNDAWAKDLKLADFALLCLDGKRKPVTEARSCHLAMAPNHAVVSRMDKVER
                LKQVLLHQQAKFGRNGSDCPDKFCLFQSETKNLLFNDNTECLARLHGKTTYEKYLGPQ
                YVAGITNLKKCSTSPLLEACEFLRK"
SE COUNT    41194 a  30122 c  32403 g  39349 t
IGIN

Query Match             96.1%;  Score 39.4;  DB 9;  Length 143068;
Best Local Similarity   97.6%;  Pred. No. 3e-05;
Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||||||||||||||||||| ||||||||||||||||||||
      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743


SULT 5
098613
CUS       AC098613              185437 bp    DNA     linear   PRI 01-AUG-2002
FINITION  Homo sapiens chromosome 3 clone RP11-24F11, complete sequence.
CESSION   AC098613
RSION     AC098613.2  GI:22038607
YWORDS    HTG.
URCE      human.
ORGANISM  Homo sapiens
          Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 185437)
AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
          Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
          and Haugen,E.D.
TITLE     Direct Submission
JOURNAL   Unpublished
FERENCE   2  (bases 1 to 185437)
AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE     Direct Submission
JOURNAL   Submitted (26-OCT-2001) Genome Center, University of Washington,
          Box 352145, Seattle, WA 98195, USA
FERENCE   3  (bases 1 to 185437)
AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
          Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
          and Haugen,E.D.
TITLE     Direct Submission
JOURNAL   Submitted (01-AUG-2002) Genome Center, University of Washington,
          Box 352145, Seattle, WA 98195, USA
MMENT     On Aug 1, 2002 this sequence version replaced gi:16445164.
          -------------- Genome Center
              Center: University of Washington Genome Center
              Center Code: UWGC
              Web site: http://www.genome.washington.edu
              Contact: uwgchtgs@u.washington.edu
          -------------- Project Information
              Center project name: chr-3
              Center clone name: RP11-24F11 (bc0137)
          ------------- Summary Statistics
              Sequencing vector: plasmid; L08752; 100% of reads
              Chemistry: Dye-terminator ET; 93% of reads
              Chemistry: Dye-terminator Big Dye; 7% of reads
              Assembly program: Phrap; version 0.990319
              Consensus quality: 184860 bases at least Q40
              Consensus quality: 185398 bases at least Q30
              Consensus quality: 185435 bases at least Q20
              Insert size: 185437; sum-of-contigs
              Quality coverage: 7.6x in Q20 bases; sum-of-contigs
          --------------.
          Overlapping Sequences:
              5':  BAC-110P12 U95626, 111014-bp overlap
              3':  RP11-509I21 (UWGC:bc0454) AC104304, 61294-bp overlap
          --------------------
          Sequence Quality Assessment:
           This entry has been annotated with sequence quality
           estimates computed by the Phrap assembly program.
           All manually edited bases have been reduced to quality zero.
           Quality levels above 40 are expected to have less than
           1 error in 10,000 bp.
           Base-by-base quality values are not generally visible from the
           GenBank flat file format but are available as part
           of this entry's ASN.1 file.

          This sequence was finished as follows unless otherwise noted:
          all regions were either double-stranded or sequenced with an
          alternate chemistry or covered by high quality data (i.e., Phred
          quality >= 30); an attempt was made to resolve all sequencing
          problems, such as compressions and repeats; all regions were
          covered by at least one plasmid subclone or more than one M13
          subclone; and the assembly was confirmed by restriction digest.
          --------------------
          Sequence Validation:
           This sequence has been validated by Multiple Complete Digest
           fingerprinting. Comparison of the experimentally derived digest
           fragments with sequence-predicted fragments is given below.
           The electronically-digested sequence consists of both insert and
           vector, in order to accurately represent the entire circular BAC.
           Small fragments below a variable cutoff (approximately 400-800 bp)
           are not resolved in the fingerprint and hence do not appear
           in the table. There are no significant remaining discrepancies
           between the experimental and predicted values. Uniquely ordered
           fragments are separated by dashed lines.
```

BglII		EcoRI		HindIII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
11786	12040	8696	8837	9531	9423
2067	2065	6	<800	6382	6501
5681	5720	1846	1824	512	<800
3716	3953	4052	3986	449	<800
875	897	1159	1130	6692	6501
910	897	54	<800	10449	10229
2169	2215	560	<800	953	1025
1705	1653	2287	2267	3124	3176
5763	5720	4905	4891	1054	1025
5844	5720	3049	3078	3985	3998
5061	5001	1377	1394	1100	1025
2625	2640	9903	9772	124	<800
725	<800	1022	1021	2671	2693
1173	1161	12606	12503	1948	1968
5747	5720	866	868	83	<800
9995	9684	9817	9772	7455	7474
999	999	3598	3601	1305	1267
4541	4503	452	<800	1047	1025
2950	2981	7549	7635	3299	3301
406	<800	2063	2075	5279	5237
3375	3602	5837	5858	6815	6926

13024	13045	13685	13328	435	<800
416	<800	4104	3986	2509	2576
10298	10140	1943	1928	2218	2172
1478	1452	3964	3986	3797	3802
287	<800	2758	2771	7758	7830
205	<800	167	<800	16192	16442
9428	9684	1307	1304	11020	10820
3809	3953	13306	13328	1978	1968
1275	1234	2423	2450	8741	8681
3351	3381	2243	2267	4795	4815
193	<800	1109	1130	14921	14985
334	<800	4131	3986	1896	1968
2546	2640	6383	6450	172	<800
9155	9107	332	<800	2091	2172
9190	9107	242	<800	2159	2172
352	<800	15571	15803	3573	3564
1157	1161	453	<800	5252	5237
3967	3953	1429	1394	4478	4428
1369	1334	655	<800	6166	6243
752	752	7881	7987	1565	1603
1624	1653	7533	7635	1645	1603
3258	3381	585	<800	573	<800
9624	9684	2998	2973	7986	7830
2841	2870	8796	8837	2313	2324
2979	3069	1331	1304	49	<800
3898	3953	7102	7161	1250	1267
4830	4804			2290	2324
2418	2459			2056	1968
6198	6066				
534	<800				
3985	3953				
1184	1161				
63	<800				

```
TURES             Location/Qualifiers
  source          1. .185437
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="3"
                  /clone="RP11-24F11"
```

```
  Query Match           96.1%;  Score 39.4;  DB 9;  Length 185437;
  Best Local Similarity  97.6%;  Pred. No. 2.9e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||| ||||||||||||||||||||
Db      10653 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 10693


RESULT 6
AF322450
LOCUS       AF322450              120793 bp    DNA     linear   HTG 06-DEC-2000
DEFINITION  Homo sapiens chromosome 17 clone BAC407I21 map 17p13.3, ***
            SEQUENCING IN PROGRESS ***, 28 unordered pieces.
ACCESSION   AF322450
VERSION     AF322450.1  GI:11559855
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 120793)
  AUTHORS   Zhao,X.T., He,M., Wan,D.F., Ye,Y., Qin,W.X., Huang,Y., Zuo,L. and
            Gu,J.R.
  TITLE     Gene clone on human chromosome 17p13.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 120793)
  AUTHORS   Zhao,X.T., He,M., Wan,D.F., Ye,Y., Qin,W.X., Huang,Y., Zuo,L. and
            Gu,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) National Laboratory For Oncogenes & Related
            Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
            200032, P. R. China
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      608: contig of 608 bp in length
            *                    gap of unknown length
            *      609     1282: contig of 674 bp in length
            *                    gap of unknown length
            *     1283     1733: contig of 451 bp in length
            *                    gap of unknown length
            *     1734     2435: contig of 702 bp in length
            *                    gap of unknown length
            *     2436     2895: contig of 460 bp in length
            *                    gap of unknown length
            *     2896     3643: contig of 748 bp in length
            *                    gap of unknown length
            *     3644     4149: contig of 506 bp in length
            *                    gap of unknown length
            *     4150     4908: contig of 759 bp in length
            *                    gap of unknown length
            *     4909     5741: contig of 833 bp in length
            *                    gap of unknown length
            *     5742     6777: contig of 1036 bp in length
            *                    gap of unknown length
            *     6778     7766: contig of 989 bp in length
            *                    gap of unknown length
            *     7767     9183: contig of 1417 bp in length
            *                    gap of unknown length
            *     9184     9931: contig of 748 bp in length
            *                    gap of unknown length
            *     9932    12274: contig of 2343 bp in length
            *                    gap of unknown length
            *    12275    16059: contig of 3785 bp in length
            *                    gap of unknown length
            *    16060    19969: contig of 3910 bp in length
            *                    gap of unknown length
```

```
          *    19970    22905: contig of 2936 bp in length
          *                    gap of unknown length
          *    22906    28210: contig of 5305 bp in length
          *                    gap of unknown length
          *    28211    34020: contig of 5810 bp in length
          *                    gap of unknown length
          *    34021    41704: contig of 7684 bp in length
          *                    gap of unknown length
          *    41705    48302: contig of 6598 bp in length
          *                    gap of unknown length
          *    48303    56789: contig of 8487 bp in length
          *                    gap of unknown length
          *    56790    65996: contig of 9207 bp in length
          *                    gap of unknown length
          *    65997    74434: contig of 8438 bp in length
          *                    gap of unknown length
          *    74435    83698: contig of 9264 bp in length
          *                    gap of unknown length
          *    83699    95956: contig of 12258 bp in length
          *                    gap of unknown length
          *    95957   107953: contig of 11997 bp in length
          *                    gap of unknown length
          *   107954   120793: contig of 12840 bp in length.
ATURES             Location/Qualifiers
   source          1. .120793
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="17"
                   /map="17p13.3"
                   /clone="BAC407I21"
SE COUNT    28674 a  31051 c  31727 g  29285 t     56 others
IGIN

Query Match             66.3%;  Score 27.2;  DB 2;  Length 120793;
Best Local Similarity   80.0%;  Pred. No. 2.3;
Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||||||| |||||||||| ||   |||| ||| |||| |
         49 CAGTTGCTCAGAAGCCTGAGATGAGAGGATTGCTTGAGCC 88


SULT 7
100748
CUS        AC100748              133544 bp    DNA     linear   PRI 01-JUN-2002
FINITION   Homo sapiens chromosome 17, clone CTD-2231E3, complete sequence.
CESSION    AC100748
RSION      AC100748.2  GI:21306865
YWORDS     HTG.
URCE       human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE    1  (bases 1 to 133544)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 17, clone CTD-2231E3
JOURNAL    Unpublished
FERENCE    2  (bases 1 to 133544)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
           Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
           Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
           Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
           MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
           McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
           Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
           Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
           Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
           Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
           Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE  3  (bases 1 to 133544)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
           Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
           Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
           Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
           Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
           Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
           Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
           Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
           Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
           Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Jun 1, 2002 this sequence version replaced gi:17048115.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
           -------------- Project Information
               Center project name: L17901
               Center clone name: 2231_E_3
           --------------.
FEATURES             Location/Qualifiers
     source          1. .133544
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2231E3"
                     /clone_lib="CITD1 Human BAC"
     repeat_region   2. .92
                     /rpt_family="AluY"
     repeat_region   642. .705
                     /rpt_family="L2"
     repeat_region   954. .1030
                     /rpt_family="AluSq/x"
     repeat_region   1033. .1159
                     /rpt_family="AluY"
     repeat_region   1162. .1470
                     /rpt_family="AluY"
     repeat_region   1474. .1578
                     /rpt_family="L2"
     repeat_region   1811. .2124
                     /rpt_family="AluSq"
     repeat_region   2227. .2356
                     /rpt_family="FLAM_C"
```

```
repeat_region   2359. .2664
                /rpt_family="AluY"
repeat_region   2666. .2972
                /rpt_family="AluSc"
repeat_region   3119. .3305
                /rpt_family="GA-rich"
repeat_region   complement(3411. .3646)
                /rpt_family="AluSx"
repeat_region   3707. .3742
                /rpt_family="(TAA)n"
repeat_region   complement(3747. .3951)
                /rpt_family="MIR"
repeat_region   4038. .4315
                /rpt_family="L1MC5"
repeat_region   4317. .4343
                /rpt_family="AT_rich"
repeat_region   complement(4344. .4729)
                /rpt_family="L1PA5"
repeat_region   complement(4730. .5037)
                /rpt_family="AluSc"
repeat_region   complement(5038. .5114)
                /rpt_family="L1PA5"
repeat_region   complement(5121. .5409)
                /rpt_family="AluSx"
repeat_region   5530. .6202
                /rpt_family="L2"
repeat_region   complement(6244. .6561)
                /rpt_family="AluSx"
repeat_region   complement(6571. .6882)
                /rpt_family="AluSq"
repeat_region   7169. .7189
                /rpt_family="GC_rich"
repeat_region   7208. .7292
                /rpt_family="(CGG)n"
repeat_region   7309. .7392
                /rpt_family="(CGGG)n"
repeat_region   7488. .7672
                /rpt_family="(CGGGG)n"
repeat_region   8823. .8853
                /rpt_family="(TG)n"
unsure          11173. .11391
                /note="single clone coverage"
repeat_region   11182. .11203
                /rpt_family="GC_rich"
unsure          11191. .11196
                /note="<30 qual single clone coverage"
unsure          11210. .11217
                /note="<30 qual single clone coverage"
unsure          11342. .11347
                /note="<30 qual single clone coverage"
unsure          11374. .11379
                /note="<30 qual single clone coverage"
repeat_region   complement(11725. .12024)
                /rpt_family="AluSx"
repeat_region   complement(13025. .13151)
                /rpt_family="AluJo"
repeat_region   15438. .15582
                /rpt_family="AluSg/x"
repeat_region   16432. .16574
                /rpt_family="FLAM_C"
repeat_region   complement(17703. .17810)
                /rpt_family="L1MB3"
repeat_region   18378. .18699
                /rpt_family="AluJo"
repeat_region   19047. .19117
                /rpt_family="L2"
repeat_region   19228. .19445
                /rpt_family="AluJo"
repeat_region   19727. .19806
                /rpt_family="L2"
repeat_region   19999. .20309
                /rpt_family="AluSx"
repeat_region   20356. .20652
                /rpt_family="AluY"
repeat_region   20768. .21077
                /rpt_family="AluSg"
repeat_region   complement(21954. .22261)
                /rpt_family="AluSx"
repeat_region   complement(22299. .22581)
                /rpt_family="AluSx"
repeat_region   complement(22597. .22895)
                /rpt_family="AluJb"
repeat_region   22956. .22976
                /rpt_family="(TTTTG)n"
repeat_region   23745. .24055
                /rpt_family="AluSq"
repeat_region   complement(24948. .25089)
                /rpt_family="AluSx"
repeat_region   complement(25144. .25317)
                /rpt_family="AluSx"
repeat_region   27682. .27757
                /rpt_family="MIR"
repeat_region   complement(27758. .28055)
                /rpt_family="AluSq"
repeat_region   28056. .28168
                /rpt_family="MIR"
repeat_region   28853. .29073
                /rpt_family="AluJo"

  Query Match             66.3%;  Score 27.2;  DB 9;  Length 133544;
  Best Local Similarity   80.0%;  Pred. No. 2.3;
  Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||||||| |||||||||| ||   |||| ||| |||| |
Db     103679 CAGTTGCTCAGAAGCCTGAGATGAGAGGATTGCTTGAGCC 103718


RESULT 8
AC006405
LOCUS       AC006405              163035 bp    DNA     linear   HTG 21-JAN-1999
DEFINITION  Homo sapiens chromosome 17 clone hRPK.107_N_19 map 17, ***
            SEQUENCING IN PROGRESS ***, 12 unordered pieces.
ACCESSION   AC006405
VERSION     AC006405.1  GI:4165361
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163035)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.107_N_19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163035)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 21, 1999 this sequence version replaced gi:4159875.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
```

```
          http://ftp.genome.washington.edu/RM/RepeatMasker.html.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     8815: contig of 8815 bp in length
          *                   gap of unknown length
          *    8816    12285: contig of 3470 bp in length
          *                   gap of unknown length
          *   12286    40956: contig of 28671 bp in length
          *                   gap of unknown length
          *   40957    64393: contig of 23437 bp in length
          *                   gap of unknown length
          *   64394    78975: contig of 14582 bp in length
          *                   gap of unknown length
          *   78976    96637: contig of 17662 bp in length
          *                   gap of unknown length
          *   96638   105967: contig of 9330 bp in length
          *                   gap of unknown length
          *  105968   145832: contig of 39865 bp in length
          *                   gap of unknown length
          *  145833   147041: contig of 1209 bp in length
          *                   gap of unknown length
          *  147042   150498: contig of 3457 bp in length
          *                   gap of unknown length
          *  150499   160979: contig of 10481 bp in length
          *                   gap of unknown length
          *  160980   163035: contig of 2056 bp in length.
ATURES              Location/Qualifiers
   source           1. .163035
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="17"
                    /map="17"
                    /clone="hRPK.107_N_19"
                    /clone_lib="RPCI-11 human BAC library"
SE COUNT    38384 a  43845 c  42661 g  37996 t    149 others
IGIN

Query Match             66.3%;  Score 27.2;  DB 2;  Length 163035;
Best Local Similarity   80.0%;  Pred. No. 2.2;
Matches   32;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||||||| |||||||||| ||   |||| ||| |||| |
      97969 CAGTTGCTCAGAAGCCTGAGATGAGAGGATTGCTTGAGCC 98008


SULT 9
136971
CUS       AL136971               82601 bp    DNA     linear   PRI 01-AUG-2000
FINITION  Human DNA sequence from clone RP3-322A2 on chromosome 6 Contains
          GSSs, complete sequence.
CESSION   AL136971
RSION     AL136971.7  GI:8670910
YWORDS    HTG.
URCE      human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE  1  (bases 1 to 82601)
AUTHORS   Harley,J.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequest@sanger.ac.uk
MMENT     On Jun 22, 2000 this sequence version replaced gi:8670574.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence has been finished according to sequence map criteria
          as follows. An attempt is made to resolve all sequencing problems,
          such as compressions and repeats, but not necessarily within known
          annotated human repeat sequence elements (e.g. Alu). Where the
          sequence is ambiguous, there is an annotation using the 'unsure'
          feature key.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr6
          RP3-322A2 is from the library RPCI-3 constructed at the Roswell
          Park Cancer Institute by the group of Pieter de Jong. For further
          details see http://bacpac.med.buffalo.edu/
          VECTOR: pCYPAC2
          This sequence is the entire insert of clone RP3-322A2 The true
          right end of clone RP11-11D8 is at 72842 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .82601
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-322A2"
                     /clone_lib="RPCI-3"
     repeat_region   436. .509
                     /note="2 copies 37 mer 98% conserved"
     repeat_region   935. .1106
                     /note="L1MB3 repeat: matches 5843. .6016 of consensus"
     repeat_region   1072. .1155
                     /note="L1MD3 repeat: matches 7654. .7740 of consensus"
     repeat_region   2213. .2444
                     /note="AluJb repeat: matches 75. .302 of consensus"
     repeat_region   2592. .2948
                     /note="MLT1I repeat: matches 62. .396 of consensus"
     repeat_region   3071. .3361
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   3362. .3429
                     /note="34 copies 2 mer ta 73% conserved"
     repeat_region   4115. .4415
                     /note="AluSx repeat: matches 1. .303 of consensus"
     repeat_region   4468. .5043
                     /note="MER34 repeat: matches 6. .545 of consensus"
     repeat_region   5062. .5239
                     /note="FAM repeat: matches 1. .172 of consensus"
     repeat_region   5302. .5424
                     /note="MER58B repeat: matches 204. .326 of consensus"
     repeat_region   5576. .5871
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   5977. .6039
                     /note="MER4D repeat: matches 920. .981 of consensus"
     repeat_region   6040. .6350
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   6351. .6409
                     /note="MER4D repeat: matches 863. .920 of consensus"
     repeat_region   6410. .6714
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   6715. .6760
                     /note="MER4D repeat: matches 819. .863 of consensus"
     repeat_region   6761. .7076
                     /note="AluJb repeat: matches 1. .303 of consensus"
     repeat_region   7077. .7522
                     /note="MER4D repeat: matches 359. .819 of consensus"
     repeat_region   7522. .7750
                     /note="MER4D repeat: matches 3. .234 of consensus"
     repeat_region   7757. .7868
```

```
                     /note="MER5A repeat: matches 25. .161 of consensus"
     repeat_region   7893. .8107
                     /note="AluJb repeat: matches 85. .293 of consensus"
     repeat_region   9105. .9398
                     /note="AluSq repeat: matches 1. .294 of consensus"
     repeat_region   9476. .9679
                     /note="TIGGER1 repeat: matches 26. .235 of consensus"
     repeat_region   9680. .9856
                     /note="AluSx repeat: matches 134. .310 of consensus"
     repeat_region   9857. .10162
                     /note="AluSc repeat: matches 1. .307 of consensus"
     repeat_region   10163. .10296
                     /note="AluSx repeat: matches 1. .134 of consensus"
     repeat_region   10297. .10489
                     /note="TIGGER1 repeat: matches 235. .400 of consensus"
     repeat_region   10490. .10801
                     /note="AluSq repeat: matches 1. .311 of consensus"
     repeat_region   10802. .11019
                     /note="TIGGER1 repeat: matches 400. .623 of consensus"
     repeat_region   11020. .11117
                     /note="AluJb repeat: matches 1. .100 of consensus"
     repeat_region   11118. .11483
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   11484. .11681
                     /note="AluJb repeat: matches 100. .290 of consensus"
     repeat_region   11691. .12100
                     /note="TIGGER1 repeat: matches 614. .1027 of consensus"
     repeat_region   12101. .12405
                     /note="AluSp repeat: matches 1. .307 of consensus"
     repeat_region   12406. .12571
                     /note="TIGGER1 repeat: matches 1027. .1193 of consensus"
     repeat_region   12572. .12871
                     /note="AluJb repeat: matches 1. .305 of consensus"
     repeat_region   12872. .13167
                     /note="TIGGER1 repeat: matches 1193. .1491 of consensus"
     repeat_region   13168. .13474
                     /note="AluY repeat: matches 1. .307 of consensus"
     repeat_region   13475. .13845
                     /note="TIGGER1 repeat: matches 1491. .1845 of consensus"
     repeat_region   13848. .14172
                     /note="MER65A repeat: matches 1. .334 of consensus"
     repeat_region   14272. .14473
                     /note="TIGGER1 repeat: matches 1836. .2037 of consensus"
     repeat_region   14474. .14750
                     /note="AluSx repeat: matches 1. .303 of consensus"
     repeat_region   14751. .14930
                     /note="TIGGER1 repeat: matches 2037. .2216 of consensus"
     repeat_region   14931. .15220
                     /note="AluJo repeat: matches 1. .307 of consensus"
     repeat_region   15221. .15420
                     /note="TIGGER1 repeat: matches 2216. .2417 of consensus"
     repeat_region   15445. .15747
                     /note="AluJb repeat: matches 1. .302 of consensus"
     misc_feature    15525. .15880
                     /note="match: GSS: Em:AQ132663"
     misc_feature    15527. .15878
                     /note="match: GSS: Em:B85322"
     repeat_region   16334. .16607
                     /note="AluJo repeat: matches 1. .291 of consensus"
     repeat_region   17456. .17567
                     /note="L1ME2 repeat: matches 5976. .6083 of consensus"
     repeat_region   18008. .18173
                     /note="AluSc repeat: matches 119. .288 of consensus"
     repeat_region   18210. .18520
                     /note="AluY repeat: matches 2. .308 of consensus"
     repeat_region   19267. .19560
                     /note="AluSx repeat: matches 3. .293 of consensus"
     repeat_region   19605. .19898
                     /note="MSTA repeat: matches 120. .426 of consensus"
     repeat_region   19899. .20283
                     /note="MSTA repeat: matches 1. .388 of consensus"
     repeat_region   20284. .20558
                     /note="AluSx repeat: matches 36. .312 of consensus"
     repeat_region   20559. .20598
                     /note="MSTA repeat: matches 387. .426 of consensus"
     repeat_region   20599. .20712
                     /note="MSTA repeat: matches 10. .120 of consensus"
     repeat_region   20730. .21075
                     /note="MLT1B repeat: matches 1. .386 of consensus"
     repeat_region   21251. .21775
                     /note="MLT1F repeat: matches 1. .535 of consensus"
     repeat_region   22248. .22614
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   22714. .23256
                     /note="MLT2B repeat: matches 1. .440 of consensus"
     repeat_region   23500. .23796
                     /note="AluSp repeat: matches 1. .294 of consensus"
     repeat_region   23988. .24221
                     /note="AluJo repeat: matches 1. .231 of consensus"
     misc_feature    complement(24774. .25308)
                     /note="match: GSS: Em:AQ606839"
     misc_feature    complement(24826. .25317)
                     /note="match: GSS: Em:AQ402803"
     repeat_region   24851. .25140
                     /note="AluSx repeat: matches 15. .306 of consensus"
     misc_feature    complement(24853. .25302)
                     /note="match: GSS: Em:AQ756978"
     repeat_region   25326. .25632
                     /note="AluSx repeat: matches 1. .294 of consensus"
     misc_feature    complement(25817. .26235)
                     /note="match: GSS: Em:AQ058462"
     repeat_region   26369. .26654
                     /note="AluSx repeat: matches 1. .286 of consensus"
     repeat_region   27278. .27586
                     /note="AluJo repeat: matches 1. .309 of consensus"
     repeat_region   28001. .28269
                     /note="AluSx repeat: matches 1. .267 of consensus"
     misc_feature    complement(28818. .29223)

  Query Match             64.9%;  Score 26.6;  DB 9;  Length 82601;
  Best Local Similarity   78.0%;  Pred. No. 4.2;
  Matches   32;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||| ||||  |||| |||||||   |||||||| |||| |
Db      15611 TCAGCTGCTCGGAAGGCTGAAATGGGAGGACTGCTTGAGCC 15651


RESULT 10
HSJ344H20/c
LOCUS       HSJ344H20             142441 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-344H20 map q15-16.1, ***
            SEQUENCING IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AL132710
VERSION     AL132710.16  GI:9931123
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142441)
  AUTHORS   Brown,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2000 this sequence version replaced gi:9864412.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ344H20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
```

```
          Sequencing vector: plasmid; L08752; 100% of reads
          Chemistry: Dye-terminator Big Dye; 99% of reads
          Chemistry: Dye-primer Big Dye; 0% of reads
          Consensus quality: 135820 bases at least Q40
          Consensus quality: 138685 bases at least Q30
          Consensus quality: 140381 bases at least Q20
          Insert size: 141741; sum-of-contigs
          Insert size: 146744; 8.7% error; agarose-fp
          Quality coverage: 3.92x in Q20 bases; sum-of-contigs Quality
          coverage: 3.87x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 8 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    26234: contig of 26234 bp in length
          *   26235 26334: gap of      100 bp
          *   26335    41373: contig of 15039 bp in length
          *   41374 41473: gap of      100 bp
          *   41474    54062: contig of 12589 bp in length
          *   54063 54162: gap of      100 bp
          *   54163    96434: contig of 42272 bp in length
          *   96435 96534: gap of      100 bp
          *   96535    99851: contig of 3317 bp in length
          *   99852 99951: gap of      100 bp
          *   99952   127995: contig of 28044 bp in length
          *  127996 128095: gap of      100 bp
          *  128096   139823: contig of 11728 bp in length
          *  139824 139923: gap of      100 bp
          *  139924   142441: contig of 2518 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q15-16.1"
                     /clone="RP3-344H20"
                     /clone_lib="RPCI-3"
     misc_feature    1. .26234
                     /note="assembly_fragment:00693
                     clone_end:T7
                     vector_side:left"
     misc_feature    26335. .41373
                     /note="assembly_fragment:01560
                     fragment_chain:1"
     misc_feature    41474. .54062
                     /note="assembly_fragment:01633
                     fragment_chain:1"
     misc_feature    54163. .96434
                     /note="assembly_fragment:00897"
     misc_feature    96535. .99851
                     /note="assembly_fragment:01047"
     misc_feature    99952. .127995
                     /note="assembly_fragment:01242"
     misc_feature    128096. .139823
                     /note="assembly_fragment:01550"
     misc_feature    139924. .142441
                     /note="assembly_fragment:01559"
BASE COUNT    42250 a  29800 c  29808 g  39881 t    702 others
ORIGIN

Query Match             64.9%;  Score 26.6;  DB 2;  Length 142441;
Best Local Similarity   78.0%;  Pred. No. 4;
Matches   32;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||| |||| |||| |||||||   |||||||| |||| |
     139723 TCAGCTGCTCGGAAGGCTGAAATGGGAGGACTGCTTGAGCC 139683
```

```
RESULT 11
AC116942
LOCUS       AC116942              202750 bp    DNA     linear   HTG 03-APR-2002
DEFINITION  Pan troglodytes clone RP43-28G15, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC116942
VERSION     AC116942.1  GI:19909416
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Pan troglodytes.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 202750)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202750)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cli
                 Center clone name: 028G15
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 201539 bases at least Q40
                 Consensus quality: 201788 bases at least Q30
                 Consensus quality: 201946 bases at least Q20
                 Insert size: 173000; agarose-fp
                 Insert size: 202450; sum-of-contigs
                 Quality coverage: 11.06x in Q20 bases; agarose-fp
                 Quality coverage: 9.45x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2159: contig of 2159 bp in length
            *    2160     2259: gap of unknown length
            *    2260    36575: contig of 34316 bp in length
            *   36576    36675: gap of unknown length
            *   36676    91546: contig of 54871 bp in length
            *   91547    91646: gap of unknown length
            *   91647   202750: contig of 111104 bp in length.
FEATURES             Location/Qualifiers
     source          1. .202750
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-28G15"
                     /clone_lib="RP43"
     misc_feature    1. .2159
```

```
                /note="assembly_fragment"
    misc_feature    2260. .36575
                /note="assembly_fragment"
    misc_feature    36576. .91546
                /note="assembly_fragment
                clone_end:SP6
                vector_side:right"
    misc_feature    91647. .202750
                /note="assembly_fragment
                clone_end:T7
                vector_side:left"
BE COUNT    51528 a  47600 c  46613 g  56708 t    301 others
IGIN

Query Match             64.9%;  Score 26.6;  DB 2;  Length 202750;
Best Local Similarity   78.0%;  Pred. No. 3.8;
Matches   32;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||| | || || || ||||| ||  ||||||| ||||||
      24734 TCAGCTACTCAGGAGGCTGAAGTAGAAGGACTGCTTGAGAC 24774


SULT 12
104983/c
CUS         AC104983               59155 bp    DNA     linear   HTG 22-DEC-2001
FINITION    Homo sapiens chromosome 18 clone RP11-352C3 map 18, LOW-PASS
            SEQUENCE SAMPLING.
CESSION     AC104983
RSION       AC104983.1  GI:17977361
YWORDS      HTG; HTGS_PHASE0.
URCE        Homo sapiens.
RGANISM     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE     1  (bases 1 to 59155)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
ITLE        Homo sapiens chromosome 18, clone RP11-352C3
JOURNAL     Unpublished
FERENCE     2  (bases 1 to 59155)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
ITLE        Direct Submission
JOURNAL     Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
MMENT       All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu

---------- Project Information
    Center project name: L22883
    Center clone name: 352_C_3
----------.
* NOTE: This record contains 70 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      724: contig of 724 bp in length
*      725 824: gap of      100 bp
*      825     1582: contig of 758 bp in length
*     1583 1682: gap of      100 bp
*     1683     2409: contig of 727 bp in length
*     2410 2509: gap of      100 bp
*     2510     3235: contig of 726 bp in length
*     3236 3335: gap of      100 bp
*     3336     4090: contig of 755 bp in length
*     4091 4190: gap of      100 bp
*     4191     4955: contig of 765 bp in length
*     4956 5055: gap of      100 bp
*     5056     5780: contig of 725 bp in length
*     5781 5880: gap of      100 bp
*     5881     6643: contig of 763 bp in length
*     6644 6743: gap of      100 bp
*     6744     7496: contig of 753 bp in length
*     7497 7596: gap of      100 bp
*     7597     8348: contig of 752 bp in length
*     8349 8448: gap of      100 bp
*     8449     9217: contig of 769 bp in length
*     9218 9317: gap of      100 bp
*     9318    10093: contig of 776 bp in length
*    10094 10193: gap of      100 bp
*    10194    10924: contig of 731 bp in length
*    10925 11024: gap of      100 bp
*    11025    11793: contig of 769 bp in length
*    11794 11893: gap of      100 bp
*    11894    12661: contig of 768 bp in length
*    12662 12761: gap of      100 bp
*    12762    13525: contig of 764 bp in length
*    13526 13625: gap of      100 bp
*    13626    14347: contig of 722 bp in length
*    14348 14447: gap of      100 bp
*    14448    15172: contig of 725 bp in length
*    15173 15272: gap of      100 bp
*    15273    16007: contig of 735 bp in length
*    16008 16107: gap of      100 bp
*    16108    16855: contig of 748 bp in length
*    16856 16955: gap of      100 bp
*    16956    17711: contig of 756 bp in length
*    17712 17811: gap of      100 bp
*    17812    18568: contig of 757 bp in length
*    18569 18668: gap of      100 bp
*    18669    19421: contig of 753 bp in length
*    19422 19521: gap of      100 bp
*    19522    20229: contig of 708 bp in length
*    20230 20329: gap of      100 bp
*    20330    21061: contig of 732 bp in length
*    21062 21161: gap of      100 bp
*    21162    21909: contig of 748 bp in length
*    21910 22009: gap of      100 bp
*    22010    22773: contig of 764 bp in length
*    22774 22873: gap of      100 bp
*    22874    23623: contig of 750 bp in length
*    23624 23723: gap of      100 bp
*    23724    24489: contig of 766 bp in length
*    24490 24589: gap of      100 bp
```

```
*    24590    25354: contig of 765 bp in length
*    25355 25454: gap of      100 bp
*    25455    26174: contig of 720 bp in length
*    26175 26274: gap of      100 bp
*    26275    27043: contig of 769 bp in length
*    27044 27143: gap of      100 bp
*    27144    27900: contig of 757 bp in length
*    27901 28000: gap of      100 bp
*    28001    28753: contig of 753 bp in length
*    28754 28853: gap of      100 bp
*    28854    29572: contig of 719 bp in length
*    29573 29672: gap of      100 bp
*    29673    30419: contig of 747 bp in length
*    30420 30519: gap of      100 bp
*    30520    31262: contig of 743 bp in length
*    31263 31362: gap of      100 bp
*    31363    32052: contig of 690 bp in length
*    32053 32152: gap of      100 bp
*    32153    32909: contig of 757 bp in length
*    32910 33009: gap of      100 bp
*    33010    33736: contig of 727 bp in length
*    33737 33836: gap of      100 bp
*    33837    34598: contig of 762 bp in length
*    34599 34698: gap of      100 bp
*    34699    35456: contig of 758 bp in length
*    35457 35556: gap of      100 bp
*    35557    36274: contig of 718 bp in length
*    36275 36374: gap of      100 bp
*    36375    37111: contig of 737 bp in length
*    37112 37211: gap of      100 bp
*    37212    37954: contig of 743 bp in length
*    37955 38054: gap of      100 bp
*    38055    38783: contig of 729 bp in length
*    38784 38883: gap of      100 bp
*    38884    39686: contig of 803 bp in length
*    39687 39786: gap of      100 bp
*    39787    40535: contig of 749 bp in length
*    40536 40635: gap of      100 bp
*    40636    41387: contig of 752 bp in length
*    41388 41487: gap of      100 bp
*    41488    42254: contig of 767 bp in length
*    42255 42354: gap of      100 bp
*    42355    43046: contig of 692 bp in length
*    43047 43146: gap of      100 bp
*    43147    43885: contig of 739 bp in length
*    43886 43985: gap of      100 bp
*    43986    44741: contig of 756 bp in length
*    44742 44841: gap of      100 bp
*    44842    45604: contig of 763 bp in length
*    45605 45704: gap of      100 bp
*    45705    46429: contig of 725 bp in length
*    46430 46529: gap of      100 bp
*    46530    47244: contig of 715 bp in length
*    47245 47344: gap of      100 bp
*    47345    48088: contig of 744 bp in length
*    48089 48188: gap of      100 bp
*    48189    48948: contig of 760 bp in length
*    48949 49048: gap of      100 bp
*    49049    49803: contig of 755 bp in length
*    49804 49903: gap of      100 bp
*    49904    50648: contig of 745 bp in length
*    50649 50748: gap of      100 bp
*    50749    51506: contig of 758 bp in length
*    51507 51606: gap of      100 bp
*    51607    52345: contig of 739 bp in length
*    52346 52445: gap of      100 bp
*    52446    53185: contig of 740 bp in length
*    53186 53285: gap of      100 bp
*    53286    54059: contig of 774 bp in length
*    54060 54159: gap of      100 bp
*    54160    54901: contig of 742 bp in length
*    54902 55001: gap of      100 bp
*    55002    55740: contig of 739 bp in length
*    55741 55840: gap of      100 bp
*    55841    56577: contig of 737 bp in length
*    56578 56677: gap of      100 bp
*    56678    57433: contig of 756 bp in length
*    57434 57533: gap of      100 bp
*    57534    58290: contig of 757 bp in length

  Query Match            62.4%;  Score 25.6;  DB 2;  Length 59155;
  Best Local Similarity  77.5%;  Pred. No. 11;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||||| || || || |||||| | |||| |||||||| |
Db      35874 CAGTTACTCAGGAGGCTGAAACAGGAGGATTGCCTGAGCC 35835


RESULT 13
AC003086
LOCUS       AC003086              110394 bp    DNA     linear   PRI 21-DEC-1999
DEFINITION  Homo sapiens BAC clone CTB-104F4 from 7q21-q22, complete sequence.
ACCESSION   AC003086
VERSION     AC003086.1  GI:2588618
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110394)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 110394)
  AUTHORS   Bradshaw,H., Graves,T. and Biewald,T.
  TITLE     The sequence of Homo sapiens BAC clone CTB-104F4
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 110394)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 110394)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 110394)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
              Center project name: H_RG104F04
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.
```

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
Clone CTB-104F4 is from the first release of the human BAC library CITB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTB-5F13, 200 bp overlap; the clone sequenced to the right is CTA-293F11, 200 bp overlap. Actual start of this clone is at base position 78804 of CTB-5F13; actual end is at 110200 of CTB-104F4.

This clone contains polymorphic bases with CTA-293F11.

ATURES	Location/Qualifiers
source	1. .110394
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q21-q22"
	/clone="CTB-104F4"
	/clone_lib="CITB-978SK-B"
repeat_region	131. .175
	/rpt_family="L1"
repeat_region	1395. .1416
	/rpt_family="L1"
repeat_region	1466. .1756
	/rpt_family="ALU"
repeat_region	2000. .2036
	/rpt_family="L1"
repeat_region	5535. .5590
	/rpt_family="L1"
repeat_region	complement(6440. .6522)
	/rpt_family="L1"
misc_feature	6804. .7069
	/note="match to EST T81349 (NID:g704234) yd27c10.s1"
misc_feature	complement(7736. .8097)
	/note="match to EST T81524 (NID:g704531) yd27c10.r1"
repeat_region	8413. .8499
	/rpt_family="L1"
repeat_region	8686. .9935
	/rpt_family="L1"
repeat_region	complement(9944. .10235)
	/rpt_family="ALU"
repeat_region	10236. .10689
	/rpt_family="L1"
repeat_region	10690. .10814
	/rpt_family="ALU"
repeat_region	10816. .12134
	/rpt_family="L1"
repeat_region	12149. .13515
	/rpt_family="L1"
repeat_region	complement(12574. .12982)
	/rpt_family="L1"
repeat_region	complement(13467. .13560)
	/rpt_family="L1"
repeat_region	13759. .13809
	/rpt_family="L1"
repeat_region	13941. .14230
	/rpt_family="ALU"
repeat_region	14315. .14376
	/rpt_family="L1"
repeat_region	15011. .15113
	/rpt_family="L1"
repeat_region	16643. .16668
	/rpt_family="L1"
repeat_region	18883. .19174
	/rpt_family="ALU"
repeat_region	19257. .19767
	/rpt_family="L1"
repeat_region	19768. .20060
	/rpt_family="ALU"
repeat_region	20080. .20596
	/rpt_family="L1"
repeat_region	22793. .22822
	/rpt_family="L1"
repeat_region	complement(23139. .23254)
	/rpt_family="ALU"
misc_feature	23255. .23722
	/note="match to EST N32155 (NID:g1152554) yy23b12.s1"
misc_feature	23255. .23458
	/note="match to EST R19544 (NID:g773154) yg26f05.r1"
repeat_region	complement(23798. .23850)
	/rpt_family="L1"
misc_feature	complement(24191. .24304)
	/note="match to EST R43896 (NID:g821774) yg26f05.s1"
repeat_region	complement(26269. .26552)
	/rpt_family="ALU"
repeat_region	26990. .27191
	/rpt_family="L1"
repeat_region	27233. .27284
	/rpt_family="L1"
misc_feature	complement(27804. .28176)
	/note="match to EST R83407 (NID:g928284) yp82d02.r1"
gene	complement(29090. .35579)
	/gene="WUGSC:H_RG104F04.1"
CDS	complement(join(29090. .30260,35551. .35579))
	/gene="WUGSC:H_RG104F04.1"
	/note="match to Y09615 (PID:g1707507) (NID:g1707506); H_RG104F04.1"
	/codon_start=1
	/product="mitochondrial transcription terminator factor"
	/protein_id="AAB83941.1"
	/db_xref="GI:2588619"
	/translation="MQSLSLGQTSISKGLNYLTIMAPGNLWHMRNNFLFGSRCWMTRF SAENIFKSVSFRLFGVKCHNTDSEPLKNEDLLKNLLTMGVDIDMARKRQPGVFHRMIT NEQDLKMFLLSKGASKEVIASIISRYPRAITRTPENLSKRWDLWRKIVTSDLEIVNIL ERSPESFFRSNNNLNLENNIKFLYSVGLTRKCLCRLLTNAPRTFSNSLDLNKQMVEFL QAAGLSLGHNDPADFVRKIIFKNPFILIQSTKRVKANIEFLRSTFNLNSEELLVLICG PGAEILDLSNDYARRSYANIKEKLFSLGCTEEEVQKFVLSYPDVIFLAEKKFNDKIDC LMEENISISQIIENPRVLDSSISTLKSRIKELVNAGCNLSTLNITLLSWSKKRYEAKL KKLSRFA"
misc_feature	complement(29284. .29492)
	/gene="WUGSC:H_RG104F04.1"
	/note="match to EST R45772 (NID:g804496)"
misc_feature	complement(30026. .30266)
	/gene="WUGSC:H_RG104F04.1"
	/note="match to EST D60352 (NID:g961991)"
repeat_region	complement(30754. .31049)
	/rpt_family="ALU"
repeat_region	34049. .34338
	/rpt_family="ALU"
repeat_region	34376. .34648
	/rpt_family="ALU"
repeat_region	34693. .34795
	/rpt_family="L1"
repeat_region	37172. .37462
	/rpt_family="ALU"
repeat_region	complement(37511. .37826)
	/rpt_family="ALU"
repeat_region	38964. .39005
	/rpt_family="ALU"
repeat_region	40696. .40733
	/rpt_family="L1"

```
Query Match             62.4%;  Score 25.6;  DB 9;  Length 110394;
Best Local Similarity   77.5%;  Pred. No. 10;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            ||| |||| ||||| |||| || ||||||||| |||| |
      34215 CAGCTGCTCAGAAGGCTGAGGTAGGAGGACTGCTTGAGCC 34254


SULT 14
011248
CUS       AC011248              111312 bp    DNA     linear   PRI 28-APR-2000
FINITION  Homo sapiens 12 BAC RP11-14P21 (Roswell Park Cancer Institute Human
          BAC Library) complete sequence.
CESSION   AC011248
RSION     AC011248.8  GI:7658299
YWORDS    HTG.
URCE      Homo sapiens.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FERENCE   1  (bases 1 to 111312)
AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
          Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
          Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
          David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
          Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
          Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
          Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
          Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
          Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
          Lau,S., Leal,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,W.,
          Logan,O., Lozado,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
          Martinez,C., McLeod,M.P., Mei,G., Merscher,S., Miller,A.,
          Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
          Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
          Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
          Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M.,
          Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
          Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G.,
          Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
          Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
          Gibbs,R.
TITLE     Direct Submission
JOURNAL   Unpublished
FERENCE   2  (bases 1 to 111312)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (05-OCT-1999) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
FERENCE   3  (bases 1 to 111312)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (28-APR-2000) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
MMENT     On Apr 28, 2000 this sequence version replaced gi:6996901.
          INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
          gc-help@bcm.tmc.edu

          CLONE LENGTH: This sequence does not necessarily represent the
          entire insert of this clone. Overlapping regions of clones are only
          sequenced and submitted once, so the sequence for the remainder of
          the insert may be found in the record for the adjacent clones.
          Overlapping clones are noted at the beginning and end of the
          Features listing.

          ANNOTATION OF FEATURES:
             STSs are identified using ePCR (Genome Res. 7:541-550) searches
          of a local database that includes entries from dbSTS, GDB, and
          local mapping efforts.
             Repeats are identified using RepeatMasker (A. Smit and P. Green,
          unpublished.) for Human and Mouse sequences.
             Genes and Region of sequence similarity are identified by BLAST
          (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
          EST and cDNA sequences. Genes demonstrate at least two exons
          flanked by consensus splice sites that maintained sequence
          continuity across the splice junctions. Sequences that are not
          identical matches are annotated as similar.

          SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
          standard of double strand coverage with a minimum of 2 clones and 2
          reads with no ambiguities or 2 chemistries with a minimum of 2
          clones and 3 reads with no ambiguities. If the sequence quality for
          a region does not meet this standard, it will be indicated in the
          annotation as Low Coverage.

          QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
          standards - estimated error rate less than 1 per 10,000 bases.
          Reports of lowest quality individual bases and measures of base
          quality are listed below. Description of the metrics can be found
          at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
          html.

          QUALSTAT-REPORT----------


          ---------- Summary Statistics ------------------------------
          Contig length:                                    150286
          Phrap values in estimate:                         149476
          Average error rate (BCM-Phrap estimate):          0.000180006
          Fraction of Phrap values less than 40 :           0.0291418
          Number of consensus changing edits:               15
          Number of N's in consensus :                      1


          ---------- Consensus changing edits ------------------------
          Position      Original+Context          Edited+Context
          22441       aaagagaaga(n)gagggctcca     aaagagaaga(a)gagggctcca
          79231       agtaaggcca(n)annttttantn    agtaaggcca(t)attttttaatg
          79233       taaggccana(n)ntttantncc     taaggccata(t)ttttaatgcc
          79234       aaggccanan(n)tttantncct     aaggccatat(t)tttaatgcct
          79239       canannttta(n)tncctngtga     catattttta(a)tgcctcgtga
          79241       nannttttant(n)cctngtgaaa    tatttttaat(g)cctcgtgaaa
          79245       tttantncct(n)gtgaaaataa     tttaatgcct(c)gtgaaaataa
          79279       agagctgtga(n)gctagaaatg     agagctgtga(a)gctagaaatg
          79440       agcctggcca(n)catcgtgaaa     agcctggcca(a)catcgtgaaa
          89920       tctcaaaaaa(n)caaaaagaag     tctcaaaaaa(a)caaaaagaag
          94117       gttttcccag(n)tcctggcaca     gttttcccag(t)tcctggcaca
          95790       gcgagactcc(n)tctcaaaaaa     gcgagactcc(a)tctcaaaaaa
          111017      ttaaggttcc(n)tcagctgaac     ttaaggttcc(a)tcagctgaac
          115887      atgccatata(n)caacaaacaa     atgccatata(a)caacaaacaa
          129846      ggaggctgaa(n)cgagtggatc     ggaggctgaa(g)cgagtggatc


          --------- Distribution of Quality < 40 Bases ---------------------

                1000|                                       *   *
                 900|                                       *   *
                 800|                                       *   *
           #     700|                                       *   *
           bases 600|                                       *   *
                 500|                                       *   *
                 400|                                   *   *   *
                 300|                                   *   *   *
                 200|                           *       *   *   *
                 100|                   *       *   *   *   *   *
                   0|   *   *   *   *   *   *   *   *   *   *   *
                     ---------------------------------------------
                         5   10   15   20   25   30   35   40
                                Phrap Value Range


          ------------------------------------------------------------------
          Version: 1.01  qxfo.
```

```
ATURES             Location/Qualifiers
   source          1. .111312
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="12"
                   /clone="RP11-14P21"
   repeat_region   1. .133
                   /rpt_family="AluSg/x"
   repeat_region   167. .193
                   /rpt_family="(GAGAA)n"
   repeat_region   553. .795
                   /rpt_family="MIR"
   repeat_region   complement(824. .901)
                   /rpt_family="MIR"
   repeat_region   946. .985
                   /rpt_family="(CA)n"
   repeat_region   complement(1467. .1648)
                   /rpt_family="MLT1G"
   repeat_region   1876. .2176
                   /rpt_family="AluSq"
   repeat_region   complement(2180. .2431)
                   /rpt_family="AluSx"
   repeat_region   complement(2432. .2788)
                   /rpt_family="THE1B"
   repeat_region   complement(2789. .2835)
                   /rpt_family="AluSx"
   repeat_region   complement(2864. .3025)
                   /rpt_family="MLT1F1"
   repeat_region   3101. .3327
                   /rpt_family="MIR"
   repeat_region   complement(3383. .3455)
                   /rpt_family="L2"
   STS             3974. .4064
                   /standard_name="G07716"
                   /db_xref="dbSTS:13562"
   repeat_region   5179. .5480
                   /rpt_family="AluSp"
   repeat_region   complement(6278. .6347)
                   /rpt_family="MER4A"
   repeat_region   6348. .6648
                   /rpt_family="AluSg"
   repeat_region   complement(6649. .7184)
                   /rpt_family="MER4A"
   repeat_region   8486. .8703
                   /rpt_family="L2"
   repeat_region   8729. .9122
                   /rpt_family="L2"
   repeat_region   9213. .9497
                   /rpt_family="L2"
   repeat_region   9498. .9798
                   /rpt_family="AluY"
   repeat_region   9799. .9887
                   /rpt_family="L2"
   repeat_region   complement(10707. .12039)
                   /rpt_family="L1MA4"
   repeat_region   12047. .12155
                   /rpt_family="L2"
   repeat_region   complement(12165. .12729)
                   /rpt_family="L2"
   repeat_region   12702. .12847
                   /rpt_family="L2"
   STS             13332. .13486
                   /standard_name="Cda1fh02"
                   /db_xref="GDB:446262"
   repeat_region   13505. .13528

Query Match             62.4%;  Score 25.6;  DB 9;  Length 111312;
Best Local Similarity   77.5%;  Pred. No. 10;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            ||||| ||    || ||||| ||| ||||||||| ||||||
      70300 CAGTTACTTGAGAGGCTGAAGTACGAGGACTGCTTGAGAC 70339
```

```
RESULT 15
AC093137
LOCUS       AC093137             142082 bp    DNA     linear   HTG 12-JUN-2002
DEFINITION  Pan troglodytes clone RP43-144H14, WORKING DRAFT SEQUENCE, 4
            ordered pieces.
ACCESSION   AC093137
VERSION     AC093137.2  GI:21392484
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 142082)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142082)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 142082)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 12, 2002 this sequence version replaced gi:15148127.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: anl
                 Center clone name: 144H14

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 141565 bases at least Q40
                 Consensus quality: 141698 bases at least Q30
                 Consensus quality: 141749 bases at least Q20
                 Insert size: 113000; agarose-fp
                 Insert size: 100000; pulse-field-gel
                 Insert size: 141782; sum-of-contigs
                 Quality coverage: 13.44x in Q20 bases; agarose-fp
                 Quality coverage: 15.18x in Q20 bases; pulse-field-gel
                 Quality coverage: 10.71x in Q20 bases; sum-of-contigs
            ----------.
```

```
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1    47102: contig of 47102 bp in length
          *   47103    47202: gap of unknown length
          *   47203    86175: contig of 38973 bp in length
          *   86176    86275: gap of unknown length
          *   86276   118675: contig of 32400 bp in length
          *  118676   118775: gap of unknown length
          *  118776   142082: contig of 23307 bp in length.
ATURES             Location/Qualifiers
   source          1. .142082
                   /organism="Pan troglodytes"
                   /db_xref="taxon:9598"
                   /clone="RP43-144H14"
                   /clone_lib="RP43"
   misc_feature    1. .47102
                   /note="assembly_fragment
                   clone_end:T7
                   vector_side:left"
   misc_feature    12236. .142082
                   /note="clone overlaps with GenBank Accession Number
                   AC093138 clone RP43-154A20 (center project name anm)"
   misc_feature    47203. .86175
                   /note="assembly_fragment"
   misc_feature    86276. .118675
                   /note="assembly_fragment"
   misc_feature    118776. .142082
                   /note="assembly_fragment
                   clone_end:SP6
                   vector_side:right"
   misc_feature    132713. .142082
                   /note="clone overlaps with GenBank Accession Number
                   AC093189 clone RP43-17P3 (center project name ann)"
SE COUNT    42473 a  27490 c  27247 g  44571 t    301 others
IGIN

Query Match             62.4%;  Score 25.6;  DB 2;  Length 142082;
Best Local Similarity   77.5%;  Pred. No. 9.9;
Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            ||| |||| ||||| ||||  ||  |||||||| |||| |
      38952 CAGCTGCTCAGAAGACTGAGGTAGGAGGACTGCTTGAGCC 38991


arch completed: July 22, 2003, 13:43:36
b time : 989 secs
```

GenCore version 5.1.6

nucleic - nucleic search, using sw model

n on: July 22, 2003, 11:16:53 ; Search time 170.5 Seconds
(without alignments)
541.536 Million cell updates/sec

tle: U95626-A-AT-42723
rfect score: 41
quence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 2185239 seqs, 1125999159 residues

tal number of hits satisfying chosen parameters: 4370478

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description
1	39.4	96.1	143068	21	AAF21105	Human low adenosin
2	39.4	96.1	143068	21	AAF21272	Human low adenosin
3	39.4	96.1	143068	21	AAA34983	Human adenosine re
4	39.4	96.1	143068	21	AAA35150	Human adenosine re
5	39.4	96.1	143068	24	ABL68124	Ovary cancer relat
6	39.4	96.1	149412	21	AAA35151	Human adenosine re
7	39.4	96.1	152740	21	AAF21273	Human low adenosin
8	25.6	62.4	1624	22	ABA15771	Human nervous syst
9	25.6	62.4	1624	22	AAK73388	Human immune/haema
10	24.2	59.0	633	22	AAL14736	Human breast cance
c 11	24.2	59.0	908	23	AAS84558	DNA encoding novel
12	24	58.5	360	22	AAI82165	Human polynucleoti
c 13	24	58.5	1001	22	AAI81867	Human polynucleoti
14	24	58.5	1311	22	AAH94457	Human foetal cDNA,
15	24	58.5	3565	16	AAQ84051	Sequence encoding
c 16	24	58.5	8728	22	AAK87277	Human immune/haema
17	24	58.5	10528	24	ABL32334	Human immune syste
c 18	24	58.5	17335	23	ABK42393	Genomic sequence #
c 19	24	58.5	19882	23	ABK42394	Genomic sequence #
20	24	58.5	21982	22	AAK67644	Human immune/haema
21	23.6	57.6	604	22	AAL20925	Human breast cance
22	23.6	57.6	604	22	AAL23598	Human breast cance
23	23.4	57.1	669	24	ABL89767	Human polynucleoti
c 24	23.4	57.1	38374	24	ABN96966	Gene #3464 used to
c 25	23.4	57.1	38374	24	ABL68363	Kidney cancer rela
c 26	23.4	57.1	38374	24	ABL68364	Kidney cancer rela
c 27	23.4	57.1	38374	24	ABL68824	Kidney cancer rela
c 28	23.4	57.1	1503900	22	AAK95240	Human neuregulin-1
c 29	23.4	57.1	1503900	22	AAK96733	Human neuregulin-1
c 30	23.2	56.6	496	22	AAI86106	Human polynucleoti
31	23.2	56.6	8835	22	AAL02806	Human reproductive
32	23	56.1	605	22	AAH07386	Human cDNA clone (
33	22.6	55.1	47090	22	AAK68725	Human immune/haema
34	22.6	55.1	47090	22	AAK78219	Human immune/haema
c 35	22.6	55.1	75384	22	AAK85590	Human immune/haema
c 36	22.6	55.1	145831	24	ABL62309	Colon adenocarcino
c 37	22.6	55.1	145831	24	ABL66806	Lung cancer relate
c 38	22.6	55.1	145831	24	ABL68588	Kidney cancer rela
c 39	22.6	55.1	145831	24	ABL69213	Prostate cancer re
c 40	22.4	54.6	322	22	AAL36435	Human musculoskele
c 41	22.4	54.6	329	22	AAI91783	Human polynucleoti
c 42	22.4	54.6	349	22	ABA14274	Human nervous syst
c 43	22.4	54.6	371	22	AAL09563	Human breast cance
c 44	22.4	54.6	379	22	AAL17455	Human breast cance
45	22.4	54.6	387	22	AAK67876	Human immune/haema

ALIGNMENTS

RESULT 1
AAF21105
ID AAF21105 standard; DNA; 143068 BP.
XX
AC AAF21105;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2672.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.

(NYCE/) NYCE J W.

Nyce JW;

WPI; 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not trigger adenosine receptors during metabolism, useful e.g. for treating cancers and respiratory obstructions -

Disclosure; Page 924-957; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

```
Query Match             96.1%;  Score 39.4;  DB 21;  Length 143068;
Best Local Similarity   97.6%;  Pred. No. 1.1e-05;
Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||||||||||||||||| ||||||||||||||||||||
      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

RESULT 2
AAF21272
ID AAF21272 standard; DNA; 143068 BP.

AAF21272;

14-MAR-2001 (first entry)

Human low adenosine antisense oligonucleotide related sequence #2839.

Low adenosine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1186-1219; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

```
  Query Match             96.1%;  Score 39.4;  DB 21;  Length 143068;
  Best Local Similarity   97.6%;  Pred. No. 1.1e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||| ||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

RESULT 3
AAA34983
ID AAA34983 standard; DNA; 143068 BP.
XX
AC AAA34983;
XX
DT 28-JUL-2000 (first entry)
XX

Human adenosine receptor related polynucleotide SEQ ID NO:2672.

Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

Homo sapiens.

WO200009525-A2.

24-FEB-2000.

03-AUG-1999; 99WO-US17712.

03-AUG-1998; 98US-0095212.

(UYEC-) UNIV EAST CAROLINA.

Nyce JW;

WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

Disclosure; Page 851-882; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 143068 BP; 41194 A; 30122 C; 32402 G; 39350 T; 0 other;

Query Match 96.1%; Score 39.4; DB 21; Length 143068;
Best Local Similarity 97.6%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            |||||||||||||||||||| ||||||||||||||||||||
      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

SULT 4
A35150

ID AAA35150 standard; DNA; 143068 BP.
XX
AC AAA35150;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:24.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1106-1138; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 143068 BP; 41194 A; 30126 C; 32402 G; 39346 T; 0 other;

Query Match 96.1%; Score 39.4; DB 21; Length 143068;
Best Local Similarity 97.6%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
```

```
          ||||||||||||||||||| ||||||||||||||||||||
      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

SULT 5
L68124
ABL68124 standard; DNA; 143068 BP.

ABL68124;

15-MAY-2002 (first entry)

Ovary cancer related gene sequence SEQ ID NO:6461.

Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid; stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous; cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma; gene; ds.

Homo sapiens.

WO200194629-A2.

13-DEC-2001.

30-MAY-2001; 2001WO-US10838.

05-JUN-2000; 2000US-209473P.
05-JUN-2000; 2000US-209531P.
18-SEP-2000; 2000US-233133P.
18-SEP-2000; 2000US-233617P.
20-SEP-2000; 2000US-234009P.
20-SEP-2000; 2000US-234034P.
20-SEP-2000; 2000US-234052P.
22-SEP-2000; 2000US-234509P.
22-SEP-2000; 2000US-234567P.
25-SEP-2000; 2000US-234923P.
25-SEP-2000; 2000US-234924P.
25-SEP-2000; 2000US-235077P.
25-SEP-2000; 2000US-235082P.
25-SEP-2000; 2000US-235134P.
25-SEP-2000; 2000US-235280P.
26-SEP-2000; 2000US-235637P.
26-SEP-2000; 2000US-235638P.
27-SEP-2000; 2000US-235711P.
27-SEP-2000; 2000US-235720P.
27-SEP-2000; 2000US-235840P.
27-SEP-2000; 2000US-235863P.
28-SEP-2000; 2000US-236028P.
28-SEP-2000; 2000US-236032P.
28-SEP-2000; 2000US-236033P.
28-SEP-2000; 2000US-236034P.
28-SEP-2000; 2000US-236109P.
28-SEP-2000; 2000US-236111P.
29-SEP-2000; 2000US-236842P.
29-SEP-2000; 2000US-236891P.
02-OCT-2000; 2000US-237172P.
02-OCT-2000; 2000US-237173P.
02-OCT-2000; 2000US-237278P.
02-OCT-2000; 2000US-237294P.
02-OCT-2000; 2000US-237295P.
02-OCT-2000; 2000US-237316P.
03-OCT-2000; 2000US-237425P.
03-OCT-2000; 2000US-237598P.
03-OCT-2000; 2000US-237604P.
03-OCT-2000; 2000US-237606P.
03-OCT-2000; 2000US-237608P.
01-NOV-2000; 2000US-244867P.
01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.

Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 6461; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

```
  Query Match             96.1%;  Score 39.4;  DB 24;  Length 143068;
  Best Local Similarity   97.6%;  Pred. No. 1.1e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||||||||||||||||||| ||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

RESULT 6
AAA35151
ID AAA35151 standard; DNA; 149412 BP.
XX
AC AAA35151;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:25.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

Disclosure; Page 1138-1171; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 149412 BP; 43049 A; 31388 C; 33852 G; 41123 T; 0 other;

```
Query Match             96.1%;  Score 39.4;  DB 21;  Length 149412;
Best Local Similarity   97.6%;  Pred. No. 1.1e-05;
Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

        1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
          |||||||||||||||||||| ||||||||||||||||||||
    49047 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 49087
```

RESULT 7
AAF21273
AAF21273 standard; DNA; 152740 BP.

AAF21273;

14-MAR-2001 (first entry)

Human low adenosine antisense oligonucleotide related sequence #2840.

Low adenosine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

Homo sapiens.

WO200062736-A2.

26-OCT-2000.

24-MAR-2000; 2000WO-US08020.

XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1219-1254; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 152740 BP; 44169 A; 32023 C; 34549 G; 41999 T; 0 other;

```
  Query Match             96.1%;  Score 39.4;  DB 21;  Length 152740;
  Best Local Similarity   97.6%;  Pred. No. 1.1e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||||| ||||||||||||||||||||
Db      49047 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 49087
```

RESULT 8
ABA15771/c
ID ABA15771 standard; DNA; 1624 BP.
XX
AC ABA15771;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8102.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;

neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

Homo sapiens.

WO200159063-A2.

16-AUG-2001.

17-JAN-2001; 2001WO-US01334.

31-JAN-2000; 2000US-0179065.
04-FEB-2000; 2000US-0180628.
24-FEB-2000; 2000US-0184664.
02-MAR-2000; 2000US-0186350.
16-MAR-2000; 2000US-0189874.
17-MAR-2000; 2000US-0190076.
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209467.
28-JUN-2000; 2000US-0214886.
30-JUN-2000; 2000US-0215135.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216880.
11-JUL-2000; 2000US-0217487.
11-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224518.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225268.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226868.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Disclosure; SEQ ID NO 8102; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABB14678-ABB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1624 BP; 443 A; 350 C; 393 G; 438 T; 0 other;

Query Match 62.4%; Score 25.6; DB 22; Length 1624;
Best Local Similarity 77.5%; Pred. No. 1.5;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            ||| | || |||||||||||| |   |||||||| |||| |
        919 CAGCTACTCAGAAGCCTGAAGTGGGAGGACTGCTTGAGCC 880
```

RESULT 9
AAK73388
AAK73388 standard; DNA; 1624 BP.

AAK73388;

06-NOV-2001 (first entry)

Human immune/haematopoietic antigen genomic sequence SEQ ID NO:28200.

Human; immune; haematopoietic; immune/haematopoietic antigen; cancer; cytostatic; gene therapy; vaccine; metastasis; ds.

Homo sapiens.

WO200157182-A2.

09-AUG-2001.

17-JAN-2001; 2001WO-US01354.

31-JAN-2000; 2000US-0179065.
04-FEB-2000; 2000US-0180628.
24-FEB-2000; 2000US-0184664.
02-MAR-2000; 2000US-0186350.
16-MAR-2000; 2000US-0189874.
17-MAR-2000; 2000US-0190076.
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.

20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 28200; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 1624 BP; 438 A; 393 C; 350 G; 443 T; 0 other;

Query Match 62.4%; Score 25.6; DB 22; Length 1624;
Best Local Similarity 77.5%; Pred. No. 1.5;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || |||||||||| | |||||||| |||| |
Db 706 CAGCTACTCAGAAGCCTGAAGTGGGAGGACTGCTTGAGCC 745

RESULT 10
AAL14736
ID AAL14736 standard; cDNA; 633 BP.
XX
AC AAL14736;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 7193.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 1300; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 633 BP; 223 A; 92 C; 178 G; 136 T; 4 other;

```
Query Match           59.0%;  Score 24.2;  DB 22;  Length 633;
Best Local Similarity 76.3%;  Pred. No. 4.4;
Matches   29;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGA 38
            |||| |||| |||||||||| ||  || | ||| |||
        522 TCAGCTGCTCAGAAGCCTGACGTANGAGAATTGCTTGA 559
```

ULT 11
84558/c
AAS84558 standard; cDNA; 908 BP.

AAS84558;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #20362.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.
P-PSDB; ABG20371.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 1; SEQ ID No 20362; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 908 BP; 265 A; 216 C; 210 G; 217 T; 0 other;

```
Query Match           59.0%;  Score 24.2;  DB 23;  Length 908;
Best Local Similarity 78.4%;  Pred. No. 4.7;
Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 AGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAG 39
              ||| ||  || ||||||||||||||||  | ||| ||
Db        820 AGTAGCGCAGCAGCCTGAAATACCAGGGTTTCCTTAG 784
```

RESULT 12
AAI82165
ID AAI82165 standard; cDNA; 360 BP.
XX
AC AAI82165;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2225.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO02234.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 2225; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 360 BP; 110 A; 73 C; 93 G; 84 T; 0 other;

```
Query Match           58.5%;  Score 24;  DB 22;  Length 360;
Best Local Similarity 75.0%;  Pred. No. 4.7;
Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | || || || ||||| ||  |||||||| |||| |
Db         49 CAGCTACTCAGGAGGCTGAAGTAGGAGGACTGCTTGAGCC 88
```

RESULT 13

I81867/c
AAI81867 standard; cDNA; 1001 BP.

AAI81867;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 1927.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorders; arthritis; inflammation; ss.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US04927.

28-FEB-2000; 2000US-0515126.
18-MAY-2000; 2000US-0577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-514838/56.
P-PSDB; AAO01936.

Isolated nucleic acids and polypeptides, useful for preventing diagnosing and treating e.g. leukaemia, inflammation and immune disorders -

Claim 1; SEQ ID NO 1927; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AAI79941-AAI93841) and the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1001 BP; 258 A; 272 C; 239 G; 229 T; 3 other;

Query Match 58.5%; Score 24; DB 22; Length 1001;
Best Local Similarity 75.0%; Pred. No. 5.8;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||| ||   |||||||| |||| |
Db        908 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 869
```

RESULT 14
H94457
AAH94457 standard; cDNA; 1311 BP.

AAH94457;

05-OCT-2001 (first entry)

Human foetal cDNA, SEQ ID NO: 1144.

XX
KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW nootropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW gene therapy; antisense therapy; cancer; immune disorder;
KW growth disorder; osteoporosis; thrombolytic disorder;
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
PN WO200155339-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02723.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 15-SEP-2000; 2000US-0663870.
PR 06-NOV-2000; 2000US-0707351.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;
XX
DR WPI; 2001-465571/50.
DR P-PSDB; AAM06782.
XX
PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -
XX
PS Example 3; Page 642; 715pp; English.
XX
CC The invention relates to novel foetal polypeptides encoded by
CC polynucleotides comprising one of 477 sequences fully defined in the
CC specification. The foetal polynucleotides and polypeptides are
CC useful in the treatment and diagnosis of diseases such as cancers,
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC disorders, nervous system disorders and inflammation. The present
CC sequence is a full length cDNA which was assembled using expressed
CC sequence tags (ESTs) found to be expressed in human foetal tissue
CC cDNA libraries as seeds.
XX
SQ Sequence 1311 BP; 375 A; 254 C; 213 G; 469 T; 0 other;

Query Match 58.5%; Score 24; DB 22; Length 1311;
Best Local Similarity 75.0%; Pred. No. 6.2;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | || ||||| ||||||||  |||| ||| || | |
Db       1223 CAGCTACTCAGAAGGCTGAAATAAGAGGATTGCTTGGGCC 1262
```

RESULT 15
AAQ84051
ID AAQ84051 standard; DNA; 3565 BP.
XX
AC AAQ84051;
XX
DT 26-OCT-1995 (first entry)
XX
DE Sequence encoding melanoma inhibiting protein (human).
XX
KW Melanoma; inhibition; cancer; melanoma; glioblastoma; neuroblastoma;
KW small cell lung cancer; neuroectodermal tumours; immunosuppressant;
KW phytohaemagglutinin; lymphocyte; interleukin 2; IL-2; detection;
KW probe; fusion protein; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

sig_peptide	1378..1449	
	/*tag= a	
exon	1378..1504	
	/*tag= b	
	/label= Exon 1.	
intron	1505..1585	
	/*tag= c	
	/label= Intron 1.	
exon	1586..1719	
	/*tag= d	
	/label= Exon 2.	
intron	1720..2803	
	/*tag= e	
	/label= Intron 2.	
exon	2804..2914	
	/*tag= f	
	/label= Exon 3.	
intron	2915..3231	
	/*tag= g	
	/label= Intron 3.	
exon	3232..3252	
	/*tag= h	
	/label= Exon 4.	
misc_feature	2216	
	/*tag= i	
	/note= "N at this position refers to an unidentied number of nucleotides."	

WO9503328-A.

02-FEB-1995.

19-JUL-1994; 94WO-EP02369.

20-JUL-1993; 93DE-4324247.

(BOEF) BOEHRINGER MANNHEIM GMBH.

Bogdahn U, Buettner R, Kaluza B;

WPI; 1995-075191/10.
P-PSDB; AAR69811.

New melanoma inhibiting protein and related nucleic acid - vectors, transformed cells, antibodies etc., useful for treating tumours and as immunosuppressant e.g. by gene therapy

Claim 9; Page 55-58; 85pp; German.

The protein encoded by this sequence has melanoma-inhibiting activity and can be used to treat cancer (melanoma, glioblastoma, neuroblastoma, small cell lung cancer, neuroectodermal tumours) or as an immunosuppressant (it inhibits IL-2 or phytohaemagglutinin induced proliferation of peripheral blood lymphocytes). Antibodies raised against the protein can be used to detect cell producing the protein and also for protein purification. Probes derived from the DNA can be used to detect sequences encoding the protein or related proteins. The protein may be expressed as a fusion protein (conjugated with dihydrofolate reductase (DHFR)).

Sequence 3565 BP; 875 A; 810 C; 959 G; 909 T; 12 other;

Query Match 58.5%; Score 24; DB 16; Length 3565;
Best Local Similarity 75.0%; Pred. No. 7.7;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
            ||||| || || ||||||| |   |||| |||||||| |
        519 CAGTTACTCAGGAGCCTGAGGTGGGAGGATTGCCTGAGTC 558
```

arch completed: July 22, 2003, 12:24:12
b time : 175.5 secs

GenCore version 5.1.6

nucleic - nucleic search, using sw model

n on: July 22, 2003, 11:16:51 ; Search time 36 Seconds
(without alignments)
349.271 Million cell updates/sec

tle: U95626-A-AT-42723
rfect score: 41
quence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

oring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 441362 seqs, 153338381 residues

tal number of hits satisfying chosen parameters: 882724

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	% Query Match	Length	DB	ID	Description
1	24	58.5	3565	1	US-08-578-649-3	Sequence 3, Appli
2	22.4	54.6	5892	3	US-08-755-587-27	Sequence 27, Appl
3	22.4	54.6	8353	3	US-08-611-587-1	Sequence 1, Appli
4	22.4	54.6	15328	2	US-08-888-497-33	Sequence 33, Appl
5	22.4	54.6	15328	4	US-09-362-230-33	Sequence 33, Appl
6	22.4	54.6	15328	5	PCT-US94-07926-33	Sequence 33, Appl
7	22.4	54.6	87350	3	US-08-781-891-79	Sequence 79, Appl
8	22.4	54.6	87543	4	US-09-791-211-3	Sequence 3, Appli
9	21.8	53.2	3224	4	US-08-965-729A-2	Sequence 2, Appli
10	21.8	53.2	3507	2	US-08-775-009-36	Sequence 36, Appl
11	21.4	52.2	43950	4	US-09-735-934A-3	Sequence 3, Appli
12	21.2	51.7	12141	4	US-09-488-671-10	Sequence 10, Appl
13	21	51.2	1927	4	US-08-952-365-3	Sequence 3, Appli
14	21	51.2	2684	1	US-08-367-227-1	Sequence 1, Appli
15	21	51.2	2686	4	US-08-952-365-5	Sequence 5, Appli
16	20.8	50.7	75	4	US-09-461-697-151	Sequence 151, App
17	20.8	50.7	542	4	US-09-461-697-136	Sequence 136, App
18	20.8	50.7	1320	1	US-08-599-252-84	Sequence 84, Appl
19	20.8	50.7	1320	1	US-08-436-074-57	Sequence 57, Appl
20	20.8	50.7	1320	5	PCT-US96-06352-84	Sequence 84, Appl
21	20.8	50.7	1320	5	PCT-US96-06583-84	Sequence 84, Appl
22	20.8	50.7	1400	4	US-09-245-281-40	Sequence 40, Appl
23	20.8	50.7	1400	4	US-09-207-359B-40	Sequence 40, Appl
24	20.8	50.7	5162	2	US-08-916-917-13	Sequence 13, Appl
25	20.8	50.7	5162	3	US-09-225-170-13	Sequence 13, Appl
26	20.8	50.7	8133	4	US-09-659-791A-10	Sequence 10, Appl
27	20.8	50.7	32042	4	US-09-245-281-44	Sequence 44, Appl
28	20.8	50.7	36651	4	US-09-738-894A-3	Sequence 3, Appli
29	20.8	50.7	62804	4	US-09-800-960-3	Sequence 3, Appli
30	20.8	50.7	162450	4	US-09-345-882-1	Sequence 1, Appli
31	20.8	50.7	176373	3	US-09-128-155-17	Sequence 17, Appl
32	20.6	50.2	3441	4	US-09-026-033-17	Sequence 17, Appl
33	20.4	49.8	45716	4	US-08-965-048-5	Sequence 5, Appli
34	20.4	49.8	45989	4	US-08-965-048-6	Sequence 6, Appli
35	20.2	49.3	393	4	US-09-574-141A-97	Sequence 97, Appl
36	20.2	49.3	399	4	US-09-574-141A-94	Sequence 94, Appl
37	20.2	49.3	848	4	US-08-905-223-27	Sequence 27, Appl
38	20.2	49.3	848	4	US-09-247-155-27	Sequence 27, Appl
39	20.2	49.3	862	4	US-09-289-349-6	Sequence 6, Appli
C 40	20.2	49.3	907	3	US-09-081-320-39	Sequence 39, Appl
C 41	20.2	49.3	907	4	US-09-574-141A-39	Sequence 39, Appl
C 42	20.2	49.3	907	4	US-09-707-780-39	Sequence 39, Appl
43	20.2	49.3	6485	3	US-09-081-320-2	Sequence 2, Appli
44	20.2	49.3	6485	4	US-09-574-141A-2	Sequence 2, Appli
45	20.2	49.3	6485	4	US-09-707-780-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-578-649-3
; Sequence 3, Application US/08578649
; Patent No. 5770366
;  GENERAL INFORMATION:
;    APPLICANT:  Ulrich Bogdan
;    APPLICANT:  Reinhard Buttner
;    APPLICANT:  Brigitte Kaluza
;    TITLE OF INVENTION:  MELANOMA-INHIBITING PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/578,649
;      FILING DATE:  29-July-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 24 247.2
;      FILING DATE:  20-July-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Andrew L. Tiajoloff
;      REGISTRATION NUMBER:  31,575
;      REFERENCE/DOCKET NUMBER:  BOER 1035-PFF/ALT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3565 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  1378..1449
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1378..1504
;    FEATURE:
```

NAME/KEY: exon
LOCATION: 1586..1719
FEATURE:
NAME/KEY: exon
LOCATION: 2804..2914
FEATURE:
NAME/KEY: exon
LOCATION: 3232..3252
FEATURE:
NAME/KEY: -
LOCATION: one-of(2216)
OTHER INFORMATION: /note= "N in position 2216
OTHER INFORMATION: denotes an indefinite number ans sequence of
OTHER INFORMATION: nucleotides"
S-08-578-649-3

Query Match 58.5%; Score 24; DB 1; Length 3565;
Best Local Similarity 75.0%; Pred. No. 0.78;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
519 CAGTTACTCAGGAGCCTGAGGTGGGAGGATTGCCTGAGTC 558

ESULT 2
S-08-755-587-27/c
Sequence 27, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 5892 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

; NAME/KEY: CDS
; LOCATION: 483..5412
; FEATURE:
; NAME/KEY: exon
; LOCATION: 481..5412
US-08-755-587-27

Query Match 54.6%; Score 22.4; DB 3; Length 5892;
Best Local Similarity 72.5%; Pred. No. 4.1;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
Db 231 CAGTTACTCAGGAGGGTGAGGTGCAAGGACTGCTTGAGCC 192

RESULT 3
US-08-611-587-1
; Sequence 1, Application US/08611587
; Patent No. 6150091
; GENERAL INFORMATION:
; APPLICANT: PANDOLFO, MASSIMO
; APPLICANT: MONTERMINI, LAURA
; APPLICANT: MOLTO, MARIA D.
; APPLICANT: Koenig, Michael
; APPLICANT: Campuzano, Victoria
; APPLICANT: Cossee, Mireille
; TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P. Patent Dept.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,587
; FILING DATE: 03-MAR-1996
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Brashears-Macatee, Sarah J.
; REGISTRATION NUMBER: 38,087
; REFERENCE/DOCKET NUMBER: D-5901
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-651-5620
; TELEFAX: 713-651-5246
; TELEX: 76-2829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 9q13
; UNITS: bp
US-08-611-587-1

Query Match 54.6%; Score 22.4; DB 3; Length 8353;
Best Local Similarity 72.5%; Pred. No. 4.6;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||  |||| |||||||   | |||||| |||| |
Db       1972 CAGCTACTTGGAAGGCTGAAATGGGAAGACTGCTTGAGCC 2011


RESULT 4
US-08-888-497-33
  Sequence 33, Application US/08888497
  Patent No. 5972677
   GENERAL INFORMATION:
     APPLICANT:  Tischfield, Jay A.
     APPLICANT:  Seilhamer, Jeffrey J.
     TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
     TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
     TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
     TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
     NUMBER OF SEQUENCES:  44
     CORRESPONDENCE ADDRESS:
       ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
       ADDRESSEE:  Russell PA
       STREET:  200 East Broward Boulevard
       CITY:  Fort Lauderdale
       STATE:  FL
       COUNTRY:  USA
       ZIP:  33301
     COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/08/888,497
       FILING DATE:
       CLASSIFICATION:
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  US/08/651,405
       FILING DATE:
       APPLICATION NUMBER:  US 08/097,354
       FILING DATE:  26-JUL-1993
     ATTORNEY/AGENT INFORMATION:
       NAME:  Manso, Peter J.
       REGISTRATION NUMBER:  32,264
       REFERENCE/DOCKET NUMBER:  IN21044-5
     TELECOMMUNICATION INFORMATION:
       TELEPHONE:  305-527-2498
       TELEFAX:  305-764-4996
   INFORMATION FOR SEQ ID NO:  33:
     SEQUENCE CHARACTERISTICS:
       LENGTH:  15328 base pairs
       TYPE:  nucleic acid
       STRANDEDNESS:  single
       TOPOLOGY:  linear
     MOLECULE TYPE:  cDNA
US-08-888-497-33

  Query Match             54.6%;  Score 22.4;  DB 2;  Length 15328;
  Best Local Similarity   72.5%;  Pred. No. 5.4;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||  |||| |||||||   | |||||| |||| |
Db       8735 CAGCTACTCAGAAGGCTGAGATAGGAGGATCACCTGAGCC 8774


RESULT 5
US-09-362-230-33
; Sequence 33, Application US/09362230
; Patent No. 6352849
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/362,230
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/888,497
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2498
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-362-230-33

  Query Match             54.6%;  Score 22.4;  DB 4;  Length 15328;
  Best Local Similarity   72.5%;  Pred. No. 5.4;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||  |||| |||||||   | |||||| |||| |
Db       8735 CAGCTACTCAGAAGGCTGAGATAGGAGGATCACCTGAGCC 8774


RESULT 6
PCT-US94-07926-33
; Sequence 33, Application PC/TUS9407926
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07926
FILING DATE: 15-JUL-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/097,354
FILING DATE: 26-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: IN21044-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305-527-2498
TELEFAX: 305-764-4996
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 15328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
CT-US94-07926-33

Query Match 54.6%; Score 22.4; DB 5; Length 15328;
Best Local Similarity 72.5%; Pred. No. 5.4;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
Db 8735 CAGCTACTCAGAAGGCTGAGATAGGAGGATCACCTGAGCC 8774

ESULT 7
S-08-781-891-79/c
Sequence 79, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
TITLE OF INVENTION: WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-79

Query Match 54.6%; Score 22.4; DB 3; Length 87350;
Best Local Similarity 72.5%; Pred. No. 8.9;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
Db 63492 CAGCTACTCAGGATGCTGAGATAAGAGGACTGCTTGAGCC 63453

RESULT 8
US-09-791-211-3/c
; Sequence 3, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 3
; LENGTH: 87543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 7421
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 7427
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 11609
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 12605
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 12742
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29370
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29422
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29979
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29980
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 29981
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30136
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 30140
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31205
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 31206
; OTHER INFORMATION: unknown
; NAME/KEY: unsure

LOCATION: 31592
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 33095
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 33160
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34066
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34072
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 36816
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 39020
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42164
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42459
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46808
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46823
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46826
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 47291
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52786
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52787
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68697
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68718
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68733
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 68739
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 69785
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 79134
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 79198
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 86336
; OTHER INFORMATION: unknown
; OTHER INFORMATION:
US-09-791-211-3

Query Match 54.6%; Score 22.4; DB 4; Length 87543;
Best Local Similarity 72.5%; Pred. No. 8.9;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || || | |||| ||| |||||||| |||| |
Db 63685 CAGCTACTCAGGATGCTGAGATAAGAGGACTGCTTGAGCC 63646

RESULT 9
US-08-965-729A-2/c
; Sequence 2, Application US/08965729A
; Patent No. 6200751
; GENERAL INFORMATION:
; APPLICANT: Jian-Ming Gu and Charles T. Esmon
; TITLE OF INVENTION: ENDOTHELIUM SPECIFIC EXPRESSION
; TITLE OF INVENTION: REGULATED BY EPCR CONTROL ELEMENTS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,729A
; FILING DATE: 07-NOV-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,718
; FILING DATE: 08-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 164 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3224 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: /note= "Nucleotides 2272 through 2729 are a
OTHER INFORMATION: thrombin responsive element"; Human
/S-08-965-729A-2

Query Match 53.2%; Score 21.8; DB 4; Length 3224;
Best Local Similarity 70.7%; Pred. No. 6.2;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
Db 198 TCAGCTACTCAGGAGGCTGAGGTAGGAGGATTGCTTGAGCC 158

RESULT 10
/S-08-775-009-36/c
Sequence 36, Application US/08775009
Patent No. 5935783
GENERAL INFORMATION:
APPLICANT: Gong, Weilong
APPLICANT: Emanuel, Beverly S.
APPLICANT: Budarf, Marcia L.
APPLICANT: Roe, Bruce
TITLE OF INVENTION: No. 5935783el Genes Mapping in the Digeorge and
TITLE OF INVENTION: Veliocardiofacial Syndrome Minimal Critical Region
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 5935783ris, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/775,009
FILING DATE: 27-DEC-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yatko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CH-0681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3507 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 163..423
FEATURE:
NAME/KEY: exon
LOCATION: 532..783
FEATURE:
NAME/KEY: exon
LOCATION: 1351..1458
FEATURE:
NAME/KEY: CDS
LOCATION: join(163..423, 532..783, 1351..1458)
/S-08-775-009-36

Query Match 53.2%; Score 21.8; DB 2; Length 3507;
Best Local Similarity 78.8%; Pred. No. 6.3;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGC 34
Db 1940 CAGGGCCTGGGAAGCCTGCAACGCCAGGACTGC 1908

RESULT 11
US-09-735-934A-3
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: LI, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

Query Match 52.2%; Score 21.4; DB 4; Length 43950;
Best Local Similarity 71.8%; Pred. No. 19;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 3 AGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
Db 36201 AGCTGCTTAGGAGGCTGAGATGGGAGGATTGCTTGAGCC 36239

RESULT 12
US-09-488-671-10/c
; Sequence 10, Application US/09488671A
; Patent No. 6187545
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Madeline M. Butler
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Lex M. Cowsert
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEPCK-CYTOSOLIC EXPRESSION
; FILE REFERENCE: RTS-0123
; CURRENT APPLICATION NUMBER: US/09/488,671A
; CURRENT FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 177
; SEQ ID NO 10
; LENGTH: 12141
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5895)...(6118)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6440)...(6621)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7206)...(7409)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7682)...(7869)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8444)...(8606)
; FEATURE:

NAME/KEY: CDS
LOCATION: (9418)...(9642)
FEATURE:
NAME/KEY: CDS
LOCATION: (9735)...(9866)
FEATURE:
NAME/KEY: CDS
LOCATION: (10502)...(10597)
FEATURE:
NAME/KEY: CDS
LOCATION: (10701)...(11155)
S-09-488-671-10

Query Match 51.7%; Score 21.2; DB 4; Length 12141;
Best Local Similarity 76.5%; Pred. No. 16;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 GCTGAGAAGCCTGAAATACCAGGACTGCCTGAGA 40
||||| | |||||| |||||||| | ||||
Db 6805 GCTGACAGGCCTGAGCCACCAGGACTTGCAGAGA 6772

ESULT 13
S-08-952-365-3
Sequence 3, Application US/08952365
Patent No. 6274311
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: A Method and Nucleotide Sequence for
Patent No. 6274311
TITLE OF INVENTION: Transforming Microorganisms
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,365
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kurdydyk, Linda M.
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 6580-63
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1927 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
IMMEDIATE SOURCE:
CLONE: EML
FEATURE:
NAME/KEY: CDS
LOCATION: 267..1832
S-08-952-365-3

Query Match 51.2%; Score 21; DB 4; Length 1927;
Best Local Similarity 73.0%; Pred. No. 11;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGA 38
|||||||| |||| || || || ||| || ||||
Db 1735 CAGTTGCTAAGAAAGCTCAAGAACAAGGTCTTACTGA 1771

RESULT 14
US-08-367-227-1
; Sequence 1, Application US/08367227
; Patent No. 5587304
; GENERAL INFORMATION:
; APPLICANT: BARRE, PIERRE
; APPLICANT: DEQUIN, SYLVIE
; APPLICANT: ANSANAY, VIRGINIE
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THE GENE FOR
; TITLE OF INVENTION: THE MALOLACTIC ENZYME OF LACTOCOCCUS LACTIS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,227
; FILING DATE: 17-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR PCT/FR94/00589
; FILING DATE: 18-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93/06003
; FILING DATE: 18-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 384-42-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2684 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 466..2085
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 392..397
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 416..421
US-08-367-227-1

Query Match 51.2%; Score 21; DB 1; Length 2684;
Best Local Similarity 73.0%; Pred. No. 13;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGA 38
|||||||| |||| || || || ||| || ||||

Db 1991 CAGTTGCTAAGAAAGCTCAAGAACAAGGTCTTACTGA 2027

RESULT 15
US-08-952-365-5
Sequence 5, Application US/08952365
Patent No. 6274311
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: A Method and Nucleotide Sequence for
Patent No. 6274311
TITLE OF INVENTION: Transforming Microorganisms
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,365
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kurdydyk, Linda M.
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 6580-63
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2686 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
IMMEDIATE SOURCE:
CLONE: mleS
FEATURE:
NAME/KEY: CDS
LOCATION: 467..2089
US-08-952-365-5

Query Match 51.2%; Score 21; DB 4; Length 2686;
Best Local Similarity 73.0%; Pred. No. 13;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGA 38
||||||||| |||| || || || ||| || |||| |
Db 1992 CAGTTGCTAAGAAAGCTCAAGAACAAGGTCTTACTGA 2028

Search completed: July 22, 2003, 11:18:24
Job time : 38 secs

GenCore version 5.1.6

DM nucleic - nucleic search, using sw model

Run on: July 22, 2003, 11:16:53 ; Search time 130 Seconds
(without alignments)
650.639 Million cell updates/sec

Title: U95626-A-AT-42723
Perfect score: 41
Sequence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39.4	96.1	143068	11	US-09-967-768A-316	Sequence 316, App
c 2	24.6	60.0	610	15	US-10-027-632-224761	Sequence 224761,
3	24	58.5	544	15	US-10-027-632-66000	Sequence 66000, A
4	24	58.5	544	15	US-10-027-632-66001	Sequence 66001, A
5	24	58.5	544	15	US-10-027-632-66002	Sequence 66002, A
6	24	58.5	544	15	US-10-027-632-298997	Sequence 298997,
7	24	58.5	544	15	US-10-027-632-298998	Sequence 298998,
8	24	58.5	544	15	US-10-027-632-298999	Sequence 298999,
9	24	58.5	1206	15	US-10-027-632-210607	Sequence 210607,
c 10	24	58.5	17335	11	US-09-764-847-1280	Sequence 1280, Ap
c 11	24	58.5	17335	15	US-10-092-154-1280	Sequence 1280, Ap
c 12	24	58.5	19882	11	US-09-764-847-1281	Sequence 1281, Ap
c 13	24	58.5	19882	15	US-10-092-154-1281	Sequence 1281, Ap
c 14	24	58.5	183337	15	US-10-020-141-5	Sequence 5, Appli
c 15	24	58.5	202001	10	US-09-734-674-3	Sequence 3, Appli
c 16	24	58.5	202001	15	US-10-274-990-3	Sequence 3, Appli
17	24	58.5	300000	15	US-10-262-552-33	Sequence 33, Appl
18	23.6	57.6	591	15	US-10-198-846-11772	Sequence 11772, A
19	23.6	57.6	654	15	US-10-027-632-234154	Sequence 234154,
20	23.4	57.1	590	15	US-10-027-632-216678	Sequence 216678,
21	23.4	57.1	644	15	US-10-027-632-254451	Sequence 254451,
22	23.4	57.1	731	15	US-10-027-632-152011	Sequence 152011,
23	23.4	57.1	2149	15	US-10-027-632-103171	Sequence 103171,
24	23.4	57.1	2149	15	US-10-027-632-103172	Sequence 103172,
c 25	23.4	57.1	38374	11	US-09-880-107-3463	Sequence 3463, Ap
c 26	23.4	57.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
c 27	23.4	57.1	1503841	10	US-09-795-686-1	Sequence 1, Appli
c 28	23.4	57.1	1503841	11	US-09-946-807-1	Sequence 1, Appli
29	23.2	56.6	8835	12	US-09-764-891-5494	Sequence 5494, Ap
30	22.8	55.6	1835	15	US-10-027-632-255911	Sequence 255911,
31	22.6	55.1	413	15	US-10-027-632-55856	Sequence 55856, A
32	22.6	55.1	413	15	US-10-027-632-301953	Sequence 301953,
33	22.6	55.1	676	15	US-10-027-632-222706	Sequence 222706,
34	22.6	55.1	676	15	US-10-027-632-222707	Sequence 222707,
c 35	22.6	55.1	879	15	US-10-027-632-2178	Sequence 2178, Ap
c 36	22.6	55.1	118067	14	US-10-081-327-32	Sequence 32, Appl
c 37	22.6	55.1	145831	11	US-09-969-708-79	Sequence 79, Appl
c 38	22.6	55.1	145831	11	US-09-954-456-2116	Sequence 2116, Ap
39	22.4	54.6	211	15	US-10-066-543-2056	Sequence 2056, Ap
c 40	22.4	54.6	322	11	US-09-764-877-2800	Sequence 2800, Ap
c 41	22.4	54.6	397	15	US-10-198-846-913	Sequence 913, App
c 42	22.4	54.6	401	15	US-10-027-632-279579	Sequence 279579,
c 43	22.4	54.6	497	15	US-10-027-632-214260	Sequence 214260,
44	22.4	54.6	507	15	US-10-027-632-206966	Sequence 206966,
45	22.4	54.6	535	15	US-10-198-846-7806	Sequence 7806, Ap

ALIGNMENTS

RESULT 1
US-09-967-768A-316
; Sequence 316, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 316
; LENGTH: 143068
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-316

Query Match 96.1%; Score 39.4; DB 11; Length 143068;
Best Local Similarity 97.6%; Pred. No. 4.4e-06;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||||||||||||||||| |||||||||||||||||||||
Db      42703 TCAGTTGCTGAGAAGCCTGACATACCAGGACTGCCTGAGAC 42743
```

RESULT 2
US-10-027-632-224761/c
; Sequence 224761, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 224761
LENGTH: 610
TYPE: DNA
ORGANISM: Human
US-10-027-632-224761

Query Match 60.0%; Score 24.6; DB 15; Length 610;
Best Local Similarity 76.9%; Pred. No. 1.9;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAG 39
|||| | || | ||| ||||||| |||||||| ||||
Db 80 TCAGCTACTCAAAAGGCTGAAATGGGAGGACTGCTTGAG 42

RESULT 3
US-10-027-632-66000
Sequence 66000, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 66000
LENGTH: 544
TYPE: DNA
ORGANISM: Human
US-10-027-632-66000

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 3.3;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || || || |||| | |||||| ||| ||||||
Db 127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166

RESULT 4
US-10-027-632-66001
; Sequence 66001, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66001
; LENGTH: 544
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-66001

Query Match 58.5%; Score 24; DB 15; Length 544;
Best Local Similarity 75.0%; Pred. No. 3.3;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || || || |||| | |||||| ||| ||||||
Db 127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166

RESULT 5
US-10-027-632-66002
; Sequence 66002, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 66002
; LENGTH: 544

```
   TYPE: DNA
   ORGANISM: Human
IS-10-027-632-66002

  Query Match             58.5%;  Score 24;  DB 15;  Length 544;
  Best Local Similarity   75.0%;  Pred. No. 3.3;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

y          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             ||| | || || || ||||  | |||||| ||| ||||||
b        127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166


IESULT 6
S-10-027-632-298997
 Sequence 298997, Application US/10027632
 GENERAL INFORMATION:
  APPLICANT: Wang, David G.
  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
  TITLE OF INVENTION:  Polymorphisms in the Human Genome
  FILE REFERENCE: 108827.129
  CURRENT APPLICATION NUMBER: US/10/027,632
  CURRENT FILING DATE:  2002-04-30
  PRIOR APPLICATION NUMBER: US 60/218,006
  PRIOR FILING DATE: 2000-07-12
  PRIOR APPLICATION NUMBER: US 60/198,676
  PRIOR FILING DATE: 2000-04-20
  PRIOR APPLICATION NUMBER: US 60/193,483
  PRIOR FILING DATE: 2000-03-29
  PRIOR APPLICATION NUMBER: US 60/185,218
  PRIOR FILING DATE: 2000-02-24
  PRIOR APPLICATION NUMBER: US 60/167,363
  PRIOR FILING DATE: 1999-11-23
  PRIOR APPLICATION NUMBER: US 60/156,358
  PRIOR FILING DATE: 1999-09-28
  PRIOR APPLICATION NUMBER: US 60/146,002
  PRIOR FILING DATE: 1999-08-09
  NUMBER OF SEQ ID NOS: 325720
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 298997
   LENGTH: 544
   TYPE: DNA
   ORGANISM: Human
IS-10-027-632-298997

  Query Match             58.5%;  Score 24;  DB 15;  Length 544;
  Best Local Similarity   75.0%;  Pred. No. 3.3;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Dy         2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             ||| | || || || ||||  | |||||| ||| ||||||
Db       127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166


IESULT 7
IS-10-027-632-298998
 Sequence 298998, Application US/10027632
 GENERAL INFORMATION:
  APPLICANT: Wang, David G.
  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
  TITLE OF INVENTION:  Polymorphisms in the Human Genome
  FILE REFERENCE: 108827.129
  CURRENT APPLICATION NUMBER: US/10/027,632
  CURRENT FILING DATE:  2002-04-30
  PRIOR APPLICATION NUMBER: US 60/218,006
  PRIOR FILING DATE: 2000-07-12
  PRIOR APPLICATION NUMBER: US 60/198,676
  PRIOR FILING DATE: 2000-04-20
  PRIOR APPLICATION NUMBER: US 60/193,483
  PRIOR FILING DATE: 2000-03-29
  PRIOR APPLICATION NUMBER: US 60/185,218
  PRIOR FILING DATE: 2000-02-24
  PRIOR APPLICATION NUMBER: US 60/167,363
  PRIOR FILING DATE: 1999-11-23
  PRIOR APPLICATION NUMBER: US 60/156,358
  PRIOR FILING DATE: 1999-09-28
  PRIOR APPLICATION NUMBER: US 60/146,002
  PRIOR FILING DATE: 1999-08-09
  NUMBER OF SEQ ID NOS: 325720
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 298998
   LENGTH: 544
   TYPE: DNA
   ORGANISM: Human
US-10-027-632-298998

  Query Match             58.5%;  Score 24;  DB 15;  Length 544;
  Best Local Similarity   75.0%;  Pred. No. 3.3;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             ||| | || || || ||||  | |||||| ||| ||||||
Db       127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166


RESULT 8
US-10-027-632-298999
 Sequence 298999, Application US/10027632
 GENERAL INFORMATION:
  APPLICANT: Wang, David G.
  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
  TITLE OF INVENTION:  Polymorphisms in the Human Genome
  FILE REFERENCE: 108827.129
  CURRENT APPLICATION NUMBER: US/10/027,632
  CURRENT FILING DATE:  2002-04-30
  PRIOR APPLICATION NUMBER: US 60/218,006
  PRIOR FILING DATE: 2000-07-12
  PRIOR APPLICATION NUMBER: US 60/198,676
  PRIOR FILING DATE: 2000-04-20
  PRIOR APPLICATION NUMBER: US 60/193,483
  PRIOR FILING DATE: 2000-03-29
  PRIOR APPLICATION NUMBER: US 60/185,218
  PRIOR FILING DATE: 2000-02-24
  PRIOR APPLICATION NUMBER: US 60/167,363
  PRIOR FILING DATE: 1999-11-23
  PRIOR APPLICATION NUMBER: US 60/156,358
  PRIOR FILING DATE: 1999-09-28
  PRIOR APPLICATION NUMBER: US 60/146,002
  PRIOR FILING DATE: 1999-08-09
  NUMBER OF SEQ ID NOS: 325720
  SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 298999
   LENGTH: 544
   TYPE: DNA
   ORGANISM: Human
US-10-027-632-298999

  Query Match             58.5%;  Score 24;  DB 15;  Length 544;
  Best Local Similarity   75.0%;  Pred. No. 3.3;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             ||| | || || || ||||  | |||||| ||| ||||||
Db       127 CAGCTACTCAGGAGGCTGAGGTGCCAGGATTGCTTGAGAC 166


RESULT 9
US-10-027-632-210607
 Sequence 210607, Application US/10027632
 GENERAL INFORMATION:
  APPLICANT: Wang, David G.
  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
  TITLE OF INVENTION:  Polymorphisms in the Human Genome
  FILE REFERENCE: 108827.129
  CURRENT APPLICATION NUMBER: US/10/027,632
  CURRENT FILING DATE:  2002-04-30
```

PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 210607
LENGTH: 1206
TYPE: DNA
ORGANISM: Human
US-10-027-632-210607

Query Match 58.5%; Score 24; DB 15; Length 1206;
Best Local Similarity 75.0%; Pred. No. 3.8;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| |||| || || ||||||| ||||| || |||| |
Db 484 CAGCTGCTCAGGAGGCTGAAATGGGAGGACGGCTTGAGCC 523

RESULT 10
US-09-764-847-1280/c
Sequence 1280, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 2003
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1280
LENGTH: 17335
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-847-1280

Query Match 58.5%; Score 24; DB 11; Length 17335;
Best Local Similarity 75.0%; Pred. No. 5.8;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || ||||| |||| | | |||||||| |||| |
Db 5266 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5227

RESULT 11
US-10-092-154-1280/c
Sequence 1280, Application US/10092154
Publication No. US20030054375A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009C1
CURRENT APPLICATION NUMBER: US/10/092,154
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2003
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1280
; LENGTH: 17335
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1280

Query Match 58.5%; Score 24; DB 15; Length 17335;
Best Local Similarity 75.0%; Pred. No. 5.8;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || ||||| |||| | | |||||||| |||| |
Db 5266 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5227

RESULT 12
US-09-764-847-1281/c
; Sequence 1281, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1281
; LENGTH: 19882
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1281

Query Match 58.5%; Score 24; DB 11; Length 19882;
Best Local Similarity 75.0%; Pred. No. 6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || ||||| |||| | | |||||||| |||| |
Db 5270 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5231

RESULT 13
US-10-092-154-1281/c
; Sequence 1281, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1281
; LENGTH: 19882
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1281

Query Match 58.5%; Score 24; DB 15; Length 19882;
Best Local Similarity 75.0%; Pred. No. 6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| | || ||||| |||| | | |||||||| |||| |
Db 5270 CAGCTACTCAGAAGGCTGAGACAGGAGGACTGCTTGAGCC 5231

RESULT 14
US-10-020-141-5/c
; Sequence 5, Application US/10020141

Publication No. US20030092013A1
GENERAL INFORMATION:
APPLICANT: McCarthy, Jeanette
APPLICANT: Ableson, Allen
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
FILE REFERENCE: MMI-002
CURRENT APPLICATION NUMBER: US/10/020,141
CURRENT FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/313,097
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: US 60/327,485
PRIOR FILING DATE: 2001-10-05
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 183337
TYPE: DNA
ORGANISM: Homo sapiens
S-10-020-141-5

Query Match 58.5%; Score 24; DB 15; Length 183337;
Best Local Similarity 75.0%; Pred. No. 8.6;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||||||| ||||| |||| | |||| ||| |||| |
Db 122236 CAGTTGCTCAGAAGGCTGAGGCAGGAGGATTGCTTGAGCC 122197

ESULT 15
IS-09-734-674-3/c
Sequence 3, Application US/09734674
Patent No. US20020081648A1
GENERAL INFORMATION:
APPLICANT: WEI, Ming-Hui et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001018
CURRENT APPLICATION NUMBER: US/09/734,674
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 202001
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(202001)
OTHER INFORMATION: n = A,T,C or G
JS-09-734-674-3

Query Match 58.5%; Score 24; DB 10; Length 202001;
Best Local Similarity 75.0%; Pred. No. 8.7;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| |||| ||||| || | || ||||| ||| |||| |
Db 88456 CAGCTGCTCAGAAGGCTAAGATGGCAGGATTGCTTGAGCC 88417

Search completed: July 22, 2003, 12:06:53
Job time : 136 secs

GenCore version 5.1.6

M nucleic - nucleic search, using sw model

un on: July 22, 2003, 11:16:52 ; Search time 1307 Seconds
(without alignments)
508.045 Million cell updates/sec

itle: U95626-A-AT-42723
erfect score: 41
equence: 1 tcagttgctgagaagcctga.....ataccaggactgcctgagac 41

coring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

earched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	26.6	64.9	352	17	B85322	B85322 RPCI11-16C1
	2	26.6	64.9	368	17	AQ132663	AQ132663 HS_3051_B
c	3	25.6	62.4	412	17	AQ181316	AQ181316 HS_3223_A
	4	25.6	62.4	471	14	BQ300489	BQ300489 PM0-KT004
c	5	25.6	62.4	509	17	AQ898508	AQ898508 HS_3127_B
	6	25.6	62.4	538	10	AW653656	AW653656 102451 MA
	7	25.6	62.4	699	10	AW978350	AW978350 EST390459
	8	25.2	61.5	453	9	AA806631	AA806631 ob71h08.s
c	9	25	61.0	326	10	AW900757	AW900757 CM1-NN100
c	10	25	61.0	457	10	AW070703	AW070703 xa29g06.x
	11	24.6	60.0	836	9	AL575170	AL575170 AL575170
	12	24	58.5	315	12	BF935457	BF935457 CM1-NT027
	13	24	58.5	337	9	AA369692	AA369692 EST81420
	14	24	58.5	374	10	AW386852	AW386852 CM0-PT004
	15	24	58.5	384	12	BF990363	BF990363 CM0-GN016
	16	24	58.5	384	12	BF995389	BF995389 CM0-GN016
	17	24	58.5	384	12	BG002012	BG002012 CM0-GN016
c	18	24	58.5	418	10	BE351027	BE351027 ht64c09.x
	19	24	58.5	423	12	BF760350	BF760350 CM1-CT060
c	20	24	58.5	430	13	BI043650	BI043650 PM4-OT020
	21	24	58.5	457	17	AQ423186	AQ423186 CITBI-E1-
	22	24	58.5	458	10	AW848768	AW848768 IL3-CT021
	23	24	58.5	470	12	BG002336	BG002336 QV4-GN025
	24	24	58.5	493	14	BQ300457	BQ300457 PM0-KT004
c	25	24	58.5	499	10	AW575409	AW575409 UI-HF-BL0
	26	24	58.5	512	12	BF858648	BF858648 RC5-FT019
	27	24	58.5	524	17	AQ784899	AQ784899 HS_3251_A
c	28	24	58.5	536	13	BM272482	BM272482 ig42d11.y
c	29	24	58.5	560	17	AQ419838	AQ419838 RPCI-11-1
	30	24	58.5	571	17	AQ504035	AQ504035 RPCI-11-2
	31	24	58.5	586	9	AL708735	AL708735 DKFZp686P
	32	24	58.5	636	17	AQ616424	AQ616424 HS_5149_A
	33	24	58.5	646	17	AG040016	AG040016 Pan trog1
	34	24	58.5	648	17	AG147167	AG147167 Pan trog1
	35	24	58.5	663	17	AG149846	AG149846 Pan trog1
	36	24	58.5	685	17	AQ075068	AQ075068 CIT-HSP-2
c	37	24	58.5	826	17	AQ749940	AQ749940 HS_5576_A
c	38	24	58.5	840	17	AQ898416	AQ898416 HS_3122_B
c	39	24	58.5	841	14	BQ719041	BQ719041 AGENCOURT
	40	23.8	58.0	1047	14	BQ049246	BQ049246 AGENCOURT
	41	23.6	57.6	184	12	BF957211	BF957211 PM1-NN120
	42	23.6	57.6	370	12	BG197441	BG197441 RST16686
c	43	23.6	57.6	409	17	B50624	B50624 CIT-HSP-523
c	44	23.6	57.6	417	17	AQ802266	AQ802266 HS_3136_B
c	45	23.6	57.6	431	9	AI039408	AI039408 ox40f02.s

ALIGNMENTS

RESULT 1
B85322
LOCUS B85322 352 bp DNA linear GSS 09-APR-1999
DEFINITION RPCI11-16C18.TP RPCI-11 Homo sapiens genomic clone RPCI-11-16C18, DNA sequence.
ACCESSION B85322
VERSION B85322.1 GI:2926454
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 352)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPCI11-16C18.TV
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from

```
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
          Research Genetics (info@resgen.com). BAC end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
          Seq primer: SP6
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /db_xref="GDB:7505825"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-16C18"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      105 a     72 c     85 g     90 t
ORIGIN

  Query Match             64.9%;  Score 26.6;  DB 17;  Length 352;
  Best Local Similarity   78.0%;  Pred. No. 14;
  Matches   32;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||| ||||  |||| |||||||   ||||||||  |||| |
Db         85 TCAGCTGCTCGGAAGGCTGAAATGGGAGGACTGCTTGAGCC 125


RESULT 2
AQ132663
LOCUS       AQ132663                368 bp    DNA     linear   GSS 23-SEP-1998
DEFINITION  HS_3051_B2_H11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3051 Col=22 Row=P, DNA sequence.
ACCESSION   AQ132663
VERSION     AQ132663.1  GI:3523389
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3051  row: P  column: 22
            Class: BAC ends
            High quality sequence stop: 368.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3051 Col=22 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      107 a     75 c     91 g     92 t      3 others
ORIGIN

  Query Match             64.9%;  Score 26.6;  DB 17;  Length 368;
  Best Local Similarity   78.0%;  Pred. No. 14;
  Matches   32;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              |||| ||||  |||| |||||||   ||||||||  |||| |
Db         87 TCAGCTGCTCGGAAGGCTGAAATGGGAGGACTGCTTGAGCC 127


RESULT 3
AQ181316/c
LOCUS       AQ181316                412 bp    DNA     linear   GSS 23-OCT-1998
DEFINITION  HS_3223_A1_F11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3223 Col=21 Row=K, DNA sequence.
ACCESSION   AQ181316
VERSION     AQ181316.1  GI:3578683
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3223  row: K  column: 21
            Class: BAC ends
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3223 Col=21 Row=K"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       97 a     95 c     75 g    144 t      1 others
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 17;  Length 412;
  Best Local Similarity   77.5%;  Pred. No. 33;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||  |||| ||||| ||  ||||||||||||| |
Db        184 CAGCTACTTGGAAGGCTGAAGTAGGAGGACTGCCTGAGCC 145


RESULT 4
BQ300489
LOCUS       BQ300489                471 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  PM0-KT0041-070201-004-b04 KT0041 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ300489
VERSION     BQ300489.1  GI:20816011
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
```

```
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
OMMENT      Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-KT0041-
            070201-004-b04&t3=2001-02-07&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 468.
EATURES                 Location/Qualifiers
     source             1. .471
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="KT0041"
                        /dev_stage="Adult"
                        /note="Organ: bladder_tumor; Vector: puc18; Site_1: SmaI;
                        Site_2: SmaI; A mini-library was made by cloning products
                        derived from  ORESTES PCR (U.S. Letters Patent application
                        No. 196,716 -  Ludwig Institute for Cancer Research)
                        profiles into  the pUC 18 vector. Reverse transcription of
                        tissue mRNA and cDNA amplification were performed under
                        low stringency conditions."
ASE COUNT       133 a    119 c    115 g    104 t
RIGIN

 Query Match             62.4%;  Score 25.6;  DB 14;  Length 471;
 Best Local Similarity   77.5%;  Pred. No. 35;
 Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

y           2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | |||||||| |||| ||   |||||||| |||| |
b         148 CAGCTACTGAGAAGGCTGACATGGTAGGACTGCTTGAGCC 187


ESULT 5
Q898508/c
OCUS        AQ898508                509 bp    DNA     linear   GSS 10-NOV-1999
EFINITION   HS_3127_B2_C10_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3127 Col=20 Row=F, DNA sequence.
CCESSION    AQ898508
ERSION      AQ898508.1  GI:6354698
EYWORDS     GSS.
OURCE       human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE    1  (bases 1 to 509)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
OMMENT      Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3127  row: F  column: 20
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 509.
FEATURES                Location/Qualifiers
     source             1. .509
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Plate=3127 Col=20 Row=F"
                        /clone_lib="CIT Approved Human Genomic Sperm Library D"
                        /sex="male"
                        /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                        E-Coli DH10B"
BASE COUNT      125 a    121 c    103 g    154 t      6 others
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 17;  Length 509;
  Best Local Similarity   77.5%;  Pred. No. 36;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||||| || ||||| ||||  || ||||| ||| |||| |
Db        237 CAGTTACTCAGAAGGCTGAGGTAGCAGGATTGCTTGAGCC 198


RESULT 6
AW653656
LOCUS       AW653656                538 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  102451 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW653656
VERSION     AW653656.1  GI:7419482
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 107  row: K  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES                Location/Qualifiers
     source             1. .538
                        /organism="Bos taurus"
                        /db_xref="taxon:9913"
                        /clone_lib="MARC 1BOV"
                        /tissue_type="pooled"
                        /lab_host="DH10B"
                        /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
```

```
                    Library made from pooled tissue from lymph node, ovary,
                    fat, hypothalamus, and pituitary."
BASE COUNT      150 a    132 c    123 g    133 t
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 10;  Length 538;
  Best Local Similarity   77.5%;  Pred. No. 36;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGA 40
              |||||||||||||||||||| | |||||  |  | | | |
Db        362 TCAGTTGCTGAGAAGCCTGACACACCAGTGCCACTTAAAA 401


RESULT 7
AW978350
LOCUS       AW978350                699 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  EST390459 MAGE resequences, MAGP Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW978350
VERSION     AW978350.1  GI:8169615
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 391
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGP"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      185 a    174 c    163 g    177 t
ORIGIN

  Query Match             62.4%;  Score 25.6;  DB 10;  Length 699;
  Best Local Similarity   77.5%;  Pred. No. 40;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | || |||||||||| |   |||||||| |||| |
Db        378 CAGCTACTCAGAAGCCTGAAGTGGGAGGACTGCTTGAGCC 417


RESULT 8
AA806631
LOCUS       AA806631                453 bp    mRNA    linear   EST 19-FEB-1998
DEFINITION  ob71h08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336863 3'
            similar to gb:X02067 H.sapiens mRNA for 7SL RNA pseudogene (HUMAN
            );contains Alu repetitive element;contains element PTR5 repetitive
            element ;, mRNA sequence.
ACCESSION   AA806631
VERSION     AA806631.1  GI:2875381
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 758    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1336863"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      105 a    107 c    116 g    125 t
ORIGIN

  Query Match             61.5%;  Score 25.2;  DB 9;  Length 453;
  Best Local Similarity   78.9%;  Pred. No. 48;
  Matches   30;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAG 39
              ||| | || |||||||||| |   |||||||| ||||
Db        386 CAGCTACTCAGAAGCCTGAAGTGGGAGGACTGCTTGAG 423


RESULT 9
AW900757/c
LOCUS       AW900757                326 bp    mRNA    linear   EST 24-MAY-2000
DEFINITION  CM1-NN1006-130400-185-c12 NN1006 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW900757
VERSION     AW900757.1  GI:8064962
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 326)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
```

```
OMMENT     Contact: Simpson A.J.G.
           Laboratory of Cancer Genetics
           Ludwig Institute for Cancer Research
           Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
           Brazil
           Tel: +55-11-2704922
           Fax: +55-11-2707001
           Email: asimpson@ludwig.org.br
           This sequence was derived from the FAPESP/LICR Human Cancer Genome
           Project. This entry can be seen in the following URL
           (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM1-NN1006-130
           400-185-c12&t3=2000-04-13&t4=1)
           Seq primer: puc 18 forward
           High quality sequence start: 12
           High quality sequence stop: 326.
EATURES             Location/Qualifiers
     source         1. .326
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_lib="NN1006"
                    /dev_stage="Adult"
                    /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                    Site_2: SmaI; A mini-library was made by cloning products
                    derived from  ORESTES PCR (U.S. Letters Patent application
                    No. 196,716 -  Ludwig Institute for Cancer Research)
                    profiles into  the pUC 18 vector. Reverse transcription of
                    tissue mRNA and cDNA amplification were performed under
                    low stringency conditions."
ASE COUNT       62 a     96 c     74 g     94 t
RIGIN

  Query Match           61.0%;  Score 25;  DB 10;  Length 326;
  Best Local Similarity   75.6%;  Pred. No. 51;
  Matches   31;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

y          1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             |||| |||| ||||| |||||  |  |||| ||| |||| |
b        213 TCAGCTGCTCAGAAGGCTGAAGCAGGAGGATTGCTTGAGCC 173


ESULT 10
W070703/c
OCUS       AW070703               457 bp    mRNA    linear   EST 13-OCT-1999
EFINITION  xa29g06.x1 NCI_CGAP_Br18 Homo sapiens cDNA clone IMAGE:2568250 3'
           similar to contains Alu repetitive element;contains element HGR
           repetitive element ;, mRNA sequence.
CCESSION   AW070703
'ERSION    AW070703.1  GI:6025701
EYWORDS    EST.
OURCE      human.
  ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
'EFERENCE  1  (bases 1 to 457)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
  JOURNAL   Unpublished (1997)
'OMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
           Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
           Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
           I.M.A.G.E. Consortium DNA Sequencing by: Washington University
           Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Seq primer: -40UP from Gibco
           High quality sequence stop: 419.
'EATURES            Location/Qualifiers
     source         1. .457
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2568250"
                    /clone_lib="NCI_CGAP_Br18"
                    /tissue_type="four pooled high-grade tumors, including two
                    primary tumors and two metastatic to ovary"
                    /lab_host="DH10B"
                    /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: SalI;
                    Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                    Library constructed by Life Technologies."
BASE COUNT       82 a    138 c    110 g    126 t      1 others
ORIGIN

  Query Match           61.0%;  Score 25;  DB 10;  Length 457;
  Best Local Similarity   75.6%;  Pred. No. 57;
  Matches   31;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             |||| |||| ||||| |||||  |  |||| ||| |||| |
Db       290 TCAGCTGCTCAGAAGGCTGAAGCAGGAGGATTGCTTGAGCC 250


RESULT 11
AL575170
LOCUS       AL575170                836 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL575170 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI063YB16 3
           prime, mRNA sequence.
ACCESSION   AL575170
VERSION     AL575170.1  GI:12936074
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES            Location/Qualifiers
     source         1. .836
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="CS0DI063YB16"
                    /clone_lib="LTI_NFL006_PL2"
                    /tissue_type="placenta"
                    /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                    was primed with a NotI-oligo(dT) primer.  Five prime end
                    enriched, double-stranded cDNA was digested with Not I and
                    cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                    vector. Library was normalized. Library was constructed by
                    Life Technologies. Contact : Feng Liang Life Technologies,
                    a division of Invitrogen 9800 Medical Center Drive
                    Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                    Email : fliang@lifetech.com URL :
                    http://fulllength.invitrogen.com"
BASE COUNT      230 a    160 c    214 g    217 t     15 others
ORIGIN

  Query Match           60.0%;  Score 24.6;  DB 9;  Length 836;
  Best Local Similarity   73.2%;  Pred. No. 98;
  Matches   30;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy         1 TCAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
             |||| ||||  | ||  ||||:  |  |||||||||||||| |
Db       587 TCAGCTGCTCGGGAGGCTGARGCAGGAGGACTGCCTGAGCC 627


RESULT 12
BF935457
LOCUS       BF935457                315 bp    mRNA    linear   EST 22-JAN-2001
```

```
DEFINITION  CM1-NT0270-271200-678-e04 NT0270 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF935457
VERSION     BF935457.1  GI:12352768
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM1&t2=CM1-NT0270-
            271200-678-e04&t3=2000-12-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 15
            High quality sequence stop: 315.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0270"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       78 a     70 c     88 g     79 t
ORIGIN

  Query Match             58.5%;  Score 24;  DB 12;  Length 315;
  Best Local Similarity   75.0%;  Pred. No. 1.2e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||| ||   |||||||| |||| |
Db        246 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 285


RESULT 13
AA369692
LOCUS       AA369692                337 bp    mRNA    linear   EST 21-APR-1997
DEFINITION  EST81420 Prostate gland I Homo sapiens cDNA 3' end similar to EST
            containing Alu repeat, mRNA sequence.
ACCESSION   AA369692
VERSION     AA369692.1  GI:2022225
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nat. Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Other_ESTs: EST81421 THC100858
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .337
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):174463"
                     /db_xref="taxon:9606"
                     /clone_lib="Prostate gland I"
                     /sex="male"
                     /dev_stage="adult, 21 yrs"
                     /note="Organ: prostate; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       76 a     79 c     97 g     84 t      1 others
ORIGIN

  Query Match             58.5%;  Score 24;  DB 9;  Length 337;
  Best Local Similarity   75.0%;  Pred. No. 1.2e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
              ||| | ||| |||| |||| ||   |||||||| |||| |
Db         94 CAGCTACTGGGAAGGCTGACATGGTAGGACTGCTTGAGCC 133


RESULT 14
AW386852
LOCUS       AW386852                374 bp    mRNA    linear   EST 04-FEB-2000
DEFINITION  CM0-PT0048-291299-141-c11 PT0048 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW386852
VERSION     AW386852.1  GI:6891511
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-PT0048-
            291299-141-c11&t3=1999-12-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 374.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="PT0048"
```

/dev_stage="Adult"
/note="Organ: pnet; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 114 a 79 c 80 g 101 t
ORIGIN

Query Match 58.5%; Score 24; DB 10; Length 374;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| |||| |||||| ||| || |||| ||| |||| |
Db 262 CAGCTGCTTAGAAGCTTGAGGCACGAGGATTGCTTGAGCC 301

RESULT 15
BF990363
LOCUS BF990363 384 bp mRNA linear EST 23-JAN-2001
DEFINITION CM0-GN0163-251000-623-a02 GN0163 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF990363
VERSION BF990363.1 GI:12396688
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-GN0163-251000-623-a02&t3=2000-10-25&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 383.
FEATURES Location/Qualifiers
source 1. .384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0163"
/dev_stage="Adult"
/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI ; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 88 a 85 c 104 g 107 t
ORIGIN

Query Match 58.5%; Score 24; DB 12; Length 384;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 CAGTTGCTGAGAAGCCTGAAATACCAGGACTGCCTGAGAC 41
||| |||| |||||| ||| || |||| ||| |||| |
Db 137 CAGCTGCTGGGAAGGCTGACATGGTAGGATTGCTTGAGCC 176

Search completed: July 22, 2003, 12:02:15
Job time : 1309 secs